Three essays on social network theory

Dissertation
submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Abel Camacho Guardian
from Mexico

approved in February 2018 at the request of
Prof. Dr. René Algesheimer
Prof. Dr. Karl Schmedders
Prof. Dr. Claudio J. Tessone

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 14.02.2018

The Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

To my parents Beatriz, Abel and my brother Arturo.

This dissertation would never have been possible without your love.

Contents

1 Introduction

A fundamental attribute of scientific hypotheses is their capability of empirical test. Thus, the use of statistical models for testing hypotheses in empirical studies has become a popular approach for analysing network data. The proliferation of network models and the increase of their complexity is partly the result of the growing number of network datasets and their use to answer social, economical and biological questions. In 2006, long after the seminal paper by Frank and Strauss (1986) introducing the first subclass of exponential random graph model (ERGM), Hunter and Handcock (2006) provided a mathematical foundation for the ERGM as an inferential tool. The use of ERGM as an inferential tool is based on mathematical results, such as the central limit theorem (CLT), which is often justified using weak-independence assumptions and large sample properties (Hayashi, 2000).

"we have said relatively little about how the number of nodes n influences the quality of the estimate $\hat{\boldsymbol{\theta}}$, even though we have relied on well-known asymptotic results about the MLE such as the use of Fisher information in approximating its covariance matrix or the implicit assumption that it is approximately normally distributed." - Hunter and Handcock (2006).

However, there is no general principle that enforces the creation of statistical network models based on the CLT. In complex systems, which network data partially covers, the attempt to understand cooperative phenomena- e.g. phase transitions, bimodal distributions- emerging as a result of assumptions at the agent-agent interaction level have been successfully addressed using tools from statistical physics (Castellano et al., 2009). In that respect, the Ising model has become one of the most important models for systems in which agents cannot be regarded as acting independently from each other.

"the import of the Ising apparatus produces a rather large spectrum of novel insights of collective social phenomena."- Galam (2008).

Yet, the same properties that are praised in the Ising model- e.g. phase transition, bimodal distributions- when modelling complex systems are downplayed when the goal of the researcher is to perform statistical analysis on network data. According to Snijders (2002), *"a bimodal distribution is very undesirable for modelling a single observation of a social network. For fitting a distribution to a single observation, the major mode of the fitted distribution should be equal, or very close, to the observed data; this is not guaranteed for families of distributions containing bimodal distributions."* .

The above argument, though, only points to the fact that statistical analysis is difficult when bimodal distributions occur and when a single network is observed. It does not explain under which assumptions at the agent-agent interaction level or assumptions at the data generating process unimodal distributions emerge. For the Ising model, unimodal distributions result by imposing assumption on the strength of the dependencies between variables (Ellis et al., 1980). However, in empirical network studies, the strength is often the problem to be addressed, and it is not controlled by researchers. Another example that the rationale of model assumptions is driven by the collected data is illustrated by Cranmer and Desmarais (2011); "The only assumptions we have made are (1) that we observe the expected values of Γ [test statistic]..." with the rationale: "Although this may seem like a strong assumption, one must keep in mind the fact that, in many cases, we will only observe a single realization of the network". For regression models, assumption (1) is a consequence of large sample properties under weak-independence assumptions, but for network data, it is unclear what the sample size is, as we discuss in Chapter 6.6.

By the time this dissertation is written a number of issues surrounding the foundation of ERGM as an inferential tool persist. First, the construction of p-values and confident interval used in empirical studies to back up conclusions rest on weak-independence assumptions which do not necessarily hold for complex data (Ellis and Newman, 1978). Second, despite the complexity of network data and the unknown large sample properties of the ERGM, the explanation for estimation problems in empirical studies is almost unanimously said to be caused by misspecified models (Robins et al., 2007; Li, 2015), which consequently implies the modification of the initial model. This contrast with estimation problems occurring in the logistic regression model, a closely relative of the ERGM, which are often said to be caused by small sample size problems (Albert and Anderson, 1984; Heinze and Schemper, 2002), and the modification of an initial model is said to cause specification bias (Zorn, 2005). The approach of modifying ERGM to solve what is called near-degeneracy problems may mask small sample size problem, as we discuss in Chapter 6.6, and it increases the chance of spurious statistical significance.

Two theoretical motivations drive our research: a) extend existing network models by introducing new mathematical tools, and b) understand how scientific evidence is constructed and justified when analysing network data and using statistical methods. For a), the starting point is to search for frameworks to generalise assumptions at the microscopic level and the macroscopic level for the ERGM, as we discuss in Chapter 2.

In Chapter 2, we introduce the class power law random graph models (PRGM) by imposing assumptions at the microscopic level. Our proposed assumptions describe how relations are being formed using the concept of q-conditional independence and q-Markov random field, where q is a parameter. We show how the parameter q in the PRGM can be interpreted as interaction terms between social mechanisms underlying the formation of networks or as an idiosyncratic correlation.

Following the work by Park and Newman (2004) for the ERGM, we show that PRGM has a macroscopic foundation via Tsallis entropy (Tsallis, 1988). Formulating a network model using Tsallis entropy is not only of mathematical interest, but it presents a future opportunity for researchers to discuss the philosophical foundation of network models based on the Boltzmann entropy. Does social network data fall in the anomaly cases where the Boltzmann-Shannon entropy fails to describe social systems? and if not, why is this the case? and how can we empirically validate it?

"we place exponential random graph models on a firm physical foundation, showing that they can be derived from first principles using maximum entropy arguments. In doing so, we argue that these models are not merely an ad hoc formulation studied primarily for their mathematical convenience, but a true and correct extension of the statistical mechanics of Boltzmann and Gibbs to the network world." - Park and Newman (2004).

Nonetheless, some caution is needed when ERGM are justified using statistical mechanics. First, statistical mechanics is concerned with properties of large systems, while ERGMs are often used to analyse systems with a few dozen or hundred of agents. Second, the statistical mechanics of Boltzmann and Gibbs deals with systems in equilibrium, and observed macroscopic properties such as entropy are assumed to be independent on the initial conditions (we discuss the equilibrium assumption for the ERGM in Chapter 6.5). Third, the additivity property of the Boltzmann-Shannon entropy is a crude approximation for finite systems or systems exhibiting long-range correlations (Tsallis, 2009), and thus network models based on non-additive entropies, e.g. PRGM, are of particular importance.

Another contribution in Chapter 2 is the construction of the q-Bernoulli random graph models and q-Markov graph models. The first class is aimed to model random networks by including solely the tendency of agents to relate with others. We show how the parameter q provides a way to increase the variability on the statistic number of relations without increasing the number of social mechanisms in the model. With the help of q-Markov graph models, we show that simple dependency structures give rise to a large class of distribution of the network statistics ranging from flat-, skewed- or bimodal distributions. Finally, using 57-friendship networks between high-school students in Switzerland and 19-friendship networks between high-school students in the USA, we show that q-Bernoulli random graph models solve the problem of placing too much probability mass on the expected value. Further, we show that the parameter q is negatively correlated with the number of agents in the networks, but positively correlated with the marginal probability of creating a relation. Although our empirical results are descriptive, they highlight the importance of developing network models placing more probability mass on the tails. A brief overview of the study presented in Chapter 2 is presented in Table 1. The proofs of results presented in Chapter 2 are given in Chapter 3.

In Chapter 4, we construct a network model for analysing multiple networks with different sizes. We show how to use fERGM to test the effects that the number of agents has on the formation of networks for two central processes, i.e. reciprocity and transitivity. We achieve this by analyzing 214 directed networks from four different datasets. Our model differs from previous approaches to analyse multiple networks by showing that our estimators are consistent as the number of observed networks tends to infinity, and we do not restrict to large networks. By changing the sample size, the validity of large sample properties is the result of assumptions on the data-generating process, which is controlled by the researcher. Although analysing multiple networks seems a straightforward approach for constructing consistent estimators, we show that the popular approach of combining two-step procedures and ERGM for analysing multiple networks do not solve the consistency issue of the ERGM, and it requires the removal of an infinite number of networks as the sample size tends to infinity. A brief overview of the study presented in Chapter 4 is presented in Table 2. The proofs of results presented in Chapter 4 are given in Chapter 5.

Motivated by the problem of understanding the underlying auxiliary assumption used in empirical network studies, in Chapter 6, we give a quick overview of the exponential random graph model. In Section 6.2, we provide a framework in which for a subclass of ERGM, termed C-ERGM, it is possible to justify statistical inference by assuming that agents are grouped in disjoint and independent communities and by observing network with a large number of communities. In Section 6.4, we review subclasses of ERGM defined by assumptions at the microscopic level, and in Section 6.5 we present necessary assumptions at the data-generating process for valid statistical inference. Further, we give a simple example showing that when the number of agents in the network tends to infinity while the number of communities is fixed, there exist cases when the CLT is violated. Our example is motivated by the mathematical results of Ellis and Newman (1978) which shows that in the Ising Courier Pott model, the central limit theorem may not hold when there are strong correlations between variables. Contrary to previous discussion affirming that estimation problems suggest misspecified models, we emphasize that estimation problems are also the result of small sample size problems, e.g. the number of communities in a network is one. In Section 6.7, we show that the concept of communities in C-ERGM shares similarities to the concept of groups in Multilevel network models.

These similarities are used to analyse Multilevel network models based on ERGM. Our main claim, here, is that each single small network should be treated as single observation and by consequence, it is not possible to perform independent statistical analysis on them, as it is done in the popular two-step procedure for multilevel network analysis (Lubbers, 2003). A brief overview of the study presented in Chapter 6 is presented in Table 3. The proofs of results presented in Chapter 6 are given in Chapter 7.

In complex systems, there is an asymmetry in arguments presented to construct scientific evidence. While the construction of analytic truth in complex systems heavily depends on mathematical results that go beyond weak-independence assumptions. The construction of factual evidence with network data often rejects the possibility of results that would imply strongly correlated variables. This rejection is necessary when doing statistical inference based on the CLT, but it does not have any behavioural foundation, and it is not clear the data-generating process needed for the CLT to hold with network data. This dissertation is a first step to solving this asymmetry of arguments, and we hope it will help social scientists, economists, policy makers to better comprehend the boundary of empirical network studies.

Table 1: Overview of the first study in the thesis.

<table>
<tr><td colspan="3">Study 1: Modelling the interdependencies between social mechanisms underlying the formation of networks</td></tr>
<tr><td>Research questions</td><td>Core contributions</td><td>Data sets</td></tr>
<tr><td>How can we model the interdependencies between social mechanisms underlying the formation of networks?
How can we increase the probability mass on the tails of network statistics?</td><td>Developing a theoretical framework to model interdependencies between social mechanisms: power law random graph model (PRGM).
Derivation of PRGM using Tsallis entropy.
Propose generalizations of three well-established network models, i.e. Bernoulli random graph model, Markov random graph model and exponential random graph model.</td><td>Our data consists of 57 friendship networks between high school students in Switzerland, and 36 friendship networks between high school students in the US.</td></tr>
<tr><td colspan="3">Main results:
-Our PRGM introduces a new parameter q, which can be interpreted as idiosyncratic correlations or an interaction term between well-defined social mechanisms.
-With the parameter q, it is possible to increase the variance of network statistics without the necessity to introduce new social mechanisms.
-Under simple dependency assumptions, network statistics are far from Gaussian, which makes it difficult to justify the construction of p-values with network data.
-The underlying assumptions of the Boltzmann-Shannon entropy are unlikely to hold in complex systems with a small number of constituents and when there are strong correlations between the actions of the constituents.
-Therefore, it is necessary to construct network models that are based on a generalisation of the Boltzmann-Shannon entropy, such as PRGM.</td></tr>
</table>

Table 2: Overview of the second study in the thesis.

<table>
<tr><td colspan="3">Study 2: Revealing the effects of network size on social mechanisms</td></tr>
<tr><td>Research questions
Does the size of the network influence the social mechanisms underlying the formation of networks?
Does the number of agents influence the probability of an agent to reciprocate a relation?</td><td>Core contributions
Developing a theoretical framework to analyze small networks, finite exponential random graph model (fERGM).
Construct tests for the effects the network size has on reciprocity and transitivity.
Show that existing frameworks for analyzing multiple networks do not have a solid mathematical foundation.</td><td>Data sets
214 directed networks from four different datasets.
84 friendship networks between students in the United States, 75 social networks of rural villages in southern India, 36 friendship networks between students in the US, and 19 friendship networks between students in Holland.</td></tr>
<tr><td colspan="3">Main results:
-We showed how our proposed fERGM can be used to analyze multiple networks.
-We showed that $\log(n)$ is a good approximation for the dependency of basic network statistics (number of links, number of reciprocated links and number of transitive triangles) on network size.
-Based on the $\log(n)$ approximation of the parameters for modeling reciprocity and transitivity, we have shown how fERGM can be used to test if these mechanisms are constant on the size of the network.</td></tr>
</table>

Table 3: Overview of the second study in the thesis.

Study 3: *On evidence and social network studies*		
Research questions	**Core contributions**	**Data sets**
Under which conditions are statistical inference based on exponential random graph valid? Why is community structure important in network models?	We review subclasses of ERGM defined by assumptions at the microscopic level. We present a subclass of ERGM with a community structure, termed C-ERGM. We give sufficient conditions on the community structure for having valid statistical inference for large networks.	Simulated data. Articles constructing subclasses of ERGM by making assumptions at the agent-agent interaction level.
Main results: -Our paper raises problems on the rationale behind p-values for ERGM based on having networks with a large number of agents but without any community structure. -Introducing community structure to ERGM helps to understand that estimation problems are not necessarily caused by misspecified models, as it is commonly assumed, but they are also the result of small sample size problem. -For the C-ERGM, we give conditions on the community structure for constructing valid p-values.		

References

Albert, A. and Anderson, J. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71(1):1–10.

Borges, E. P. (2004). A possible deformed algebra and calculus inspired in nonextensive thermostatistics. *Physica A: Statistical Mechanics and its Applications*, 340(1):95–101.

Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591.

Cranmer, S. J. and Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Political Analysis*, 19(1):66–86.

Ellis, R. S. and Newman, C. M. (1978). Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 44(2):117–139.

Ellis, R. S., Newman, C. M., and Rosen, J. S. (1980). Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 51(2):153–169.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.

Galam, S. (2008). Sociophysics: a review of galam models. *International Journal of Modern Physics C*, 19(03):409–440.

Hayashi, F. (2000). Econometrics.

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in medicine*, 21(16):2409–2419.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3).

Li, K. (2015). *Degeneracy, Duration, and Co-evolution: Extending Exponential Random Graph Models (ERGM) for Social Network Analysis*. PhD thesis.

Nivanen, L., Le Mehaute, A., and Wang, Q. (2003). Generalized algebra within a nonextensive statistics. *Reports on Mathematical Physics*, 52(3):437–444.

Park, J. and Newman, M. E. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6):066117.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2):173–191.

Sears, T. D. and Sunehag, P. (2007). Induced semantics for undirected graphs: another look at the hammersley-clifford theorem. In *BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING: 27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 954, pages 125–132. AIP Publishing.

Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models.

Tsallis, C. (1988). Possible generalization of boltzmann-gibbs statistics. *Journal of statistical physics*, 52(1-2):479–487.

Tsallis, C. (2009). Nonadditive entropy: the concept and its use. *The European Physical Journal A-Hadrons and Nuclei*, 40(3):257–266.

Van Der Pol, J. (2016). Introduction to network modeling using exponential random graph models (ergm).

Zorn, C. (2005). A solution to separation in binary response models. *Political Analysis*, 13(2):157–170.

2 Modelling the interdependencies between social mechanisms underlying the formation of networks

Abstract

During the past years, the introduction of new assumptions in network models has been mostly motivated to prove mathematical results needed for constructing p-values, and not to describe principles underlying the formation of networks that cannot be described by existing models. In this paper, we propose a network model, power law random graph model (PRGM), formulated at the agent-agent interaction level using the concept of q-conditional independence where q is a parameter. The parameter q can be interpreted as idiosyncratic correlations or interaction terms between well-defined social mechanisms, and it aims to introduce interdependencies between social mechanisms underlying the formation of networks. We construct two subclasses of PRGM that generalize well-known network models. First, we present the family of q-Bernoulli random graph model, which reduces to the classical model for $q = 1$, and we show how the parameter q can substantially increase the variance of network statistics without the necessity to introduce new social mechanisms. Second, we introduce the q-Markov random graph models, and we show how under simple assumptions the distributions of networks statistics are far from Gaussian distributions, which makes it difficult to justify the construction of p-values with network data. Next, motivated by the derivation of ERGM via the Boltzmann-Shannon entropy by Park and Newman, we present a second formulation of the PRGM via Tsallis entropy. Finally, applying the q-Bernoulli random graph models to two network datasets of friendships between students in classrooms in Switzerland and the United States, we show how q helps to address the problem of models placing too much probability mass on the expected value without adding new social mechanisms to the model.

2.1 Introduction

The inquiry of social and economic problems need to determine the truth of factual questions. How do friendships affect health behaviour (Centola, 2011; Valente et al., 2009), delinquent behaviour (Goldenberg et al., 2001; Calvó-Armengol and Zenou, 2004)? How does the structure of a social network influence the spread of diseases (Eubank et al., 2004)? Or the spread of information (Bakshy et al., 2012)? How does the financial network structure alleviate or intensifies the fragility of the financial sector (Betz et al., 2016; Haldane and May, 2011)? The above problems, among many others, occur in systems where each agent's belief, decision, and

Author Statement: This work has been done together with René Algesheimer and Claudio Tessone.

action is influenced by and influences the beliefs, decisions, and actions of the surrounded others. When decisions of agents are influenced by each other and non-independence assumptions are not satisfied in a dataset, classical statistical models are not suitable for constructing factual evidence using the dataset. This is particularly true for network data. As a result, methodologies to search for factual truth based on network data have been developed, i.e. stochastic actor oriented models (SAOM) (Snijders et al., 2010), exponential random graph models (ERGM) (Wasserman and Pattison, 1996), stochastic block models (Holland et al., 1983), among others.

The factual evidence constructed with SAOM and ERGM is supported by the statistical frameworks of p-values (De La Haye et al., 2011; Buchmann et al., 2014; Huitsing and Veenstra, 2012; Lewis, 2013), and thus the nature of the evidence lies on the auxiliary assumptions used to construct p-values. Broadly speaking, auxiliary assumptions to construct p-values are of two types (i) at the data generating process or (ii) at the agent-agent interaction level. When they are of type (i), researchers are obligated to control the data collection process to satisfy the assumptions. Independence of observations is an example of assumption of type (i) used in statistical analyses (Freedman, 1987; Hayashi, 2000) but violated in network data. When they are of type (ii), researchers are obligated to provide evidence showing the plausibility of the auxiliary assumptions for the agents and system under considerations. The Gaussian distribution of the error connecting the dependent variable and the independent variables in linear regression models is an example of an assumption of type (ii). P-values in SAOM and ERGM are constructed by approximating the distribution of network statistics, e.g. number of relations/links, by Gaussian distributions (Hunter and Handcock, 2006); and Gaussianity is conjecture to hold based on the central limit theorem (Hunter and Handcock, 2006) and computer simulations (Snijders et al., 2010) .

During the last years, researchers have introduced additional assumptions to existing net-

In (Snijders et al., 2010), it was said that SAOM are agent-based modelling but with possibility of performing statistical inference. However, it fell short in providing any mathematical justification for the p-values used in the statistical inference.

The absence of formal proofs to justify Gaussian approximations also occurs in other network models (Maslov and Sneppen, 2002; Milo et al., 2002).

work models for the central limit theorem to hold (Chandrasekhar and Jackson, 2015; Schweinberger and Handcock, 2015; Camacho Guardian, 2016). However, adding auxiliary assumptions to network models has three shortcomings: 1) it leaves researchers with the additional burden to provide evidence showing that the assumptions are met in their datasets, 2) it narrows the applicability of the model, and 3) it puts most of its probability mass on a few class of networks, e.g. networks which network statistics are approximately equal to the expected value. Thereby, our approach to constructing a new class of network models is not driven by the search for conditions for the CLT to hold, but our aim is to introduce a new formulation of network models that: 1) introduces larger probability mass on the tail of network statistics without increasing the social mechanisms in the model, and 2) adds interdependencies between social mechanisms underlying the formation of networks.

In this paper, our main contribution is introducing a new class of network models, power law random graph models (PRGM), aiming to model the interdependencies between social mechanisms underlying the formation of networks. We present two formulations for PRGM: 1) at the agent-agent interaction level and 2) at the macroscopic level using the principle of maximum entropy. The first formulation is done by generalising the concept of Markov random field, and it follows the definition of q-independence proposed by Sears and Sunehag (2007). Similar to ERGM, PRGM is defined by a set of social mechanisms but with a new parameter q, which can be interpreted as idiosyncratic correlations or as a form of interaction terms between social mechanisms.

Next, we construct the classes q-Bernoulli random graph models and q-Markov graph models. The first class is aimed to model random networks by including solely the tendency of agents to relate with others. We show how the parameter q provides a way to increase the variability on the statistic number of relations without increasing the number of social mechanisms in the model. Despite its simplicity, for certain parameters, the model exhibits a bimodal distribution, and it is caused by strong correlations between the variables. With the help of q-Markov graph models, we show that simple dependency structures give rise to a large class of distribution of the network statistics ranging from flat-, skewed- or bimodal distributions. The large possibilities of distribution emerging from simple dependency structure expose the weakness of Gaussian distribution used for statistical network analysis. These observations extend the discussion concerning false finding in the social network literature (Andriani and McKelvey, 2007, 2009; Camacho Guardian, 2016) and open new questions. How plausible are the auxiliary assumptions used for the construction of factual evidence with network data? In particular, those assumptions for justifying asymptotic distribution of the network statistics.

Next, we provide a second derivation of PRGM based on the entropy introduced by Tsallis (1988). This derivation generalises network models based on the Boltzmann-Shannon entropy (Park and Newman, 2004). In particular, we show that ERGM is a special case of PRGM. Finally, using 57-friendship networks between high-school students in Switzerland and 19-friendship networks between high-school students in the USA, we show that q-Bernoulli random graph models solve the problem of placing too much probability mass on the expected value. Further, we show that the parameter q is negatively correlated with the number of agents in the networks, but positively correlated with the marginal probability of creating a relation. Although our empirical results are descriptive, they highlight the importance of developing network models placing more probability mass on the tails.

2.2 Networks and hyponetworks

We now introduce the basic notation and concepts necessary to construct the power law random graph model, PRGM. We assume that there exists a fixed, finite set, V, composed of n agents and a set of relational variables $\boldsymbol{e}_S$ with $S = \{ij : 1 \leq i, j \leq n \text{ and } i \neq j\}$. If $e_{ij} = 1$, we say that agent i is related to agent j, otherwise $e_{ij} = 0$ and agent i is not related to agent j. Sometimes, it is necessary to express a constraint that a set of agents V' can decide having or not certain relations. This is expressed by considering a subset of variables $\boldsymbol{e}_{S'}$ where $V' = V(\boldsymbol{e}_{S'})$ and $V(\boldsymbol{e}_{S'})$ are the agents associated to the variables $\boldsymbol{e}_{S'}$. The set $\boldsymbol{e}_{S'}$ represents the set of relational variables that are updated by agents actions. For instance, $(e_{ij}) = 1$ differs from

$(e_{ij}, e_{ji}) = (1, 0)$ as in the last case agent j decides to not relate to agent i while in the first case agent j is not allowed to decide to have or not have a relation with agent i. In the second case, the decision of agent i to update its relation to j considers if agent j has decided to relate to him or not. A realisation of a set of variables $\boldsymbol{e}_{S'}$ is called an *hyponetwork*; and it is called a network when $ij \in S'$ implies that $ji \in S'$. Two hyponetworks $\bar{\boldsymbol{e}}_{S'}$, $\bar{\boldsymbol{e}}_{S''}$ are said to be isomorphic, $\bar{\boldsymbol{e}}_{S'} \sim \bar{\boldsymbol{e}}_{S''}$ if and only if there is a bijective function $f : V(\boldsymbol{e}_{S'}) \to V(\boldsymbol{e}_{S''})$ that preserves relations and the possible relations to be updated, i.e. $\bar{\boldsymbol{e}}_{ij} = \bar{\boldsymbol{e}}_{f(i)f(j)}$ and $\bar{\boldsymbol{e}}_{f^{-1}(i)f^{-1}(j)} = \bar{\boldsymbol{e}}_{ij}$. If H represents an hyponetwork and $\bar{\boldsymbol{e}}_{S'} \sim H$, we say that $\bar{\boldsymbol{e}}_{S'}$ posses structure H.

Next, let us imagine the decision of agent i to update the relation to agent j, e_{ij}, while considering some relation between certain agents, $\boldsymbol{e}_{S'} \ni e_{ij}$. This may occur if agent i can only observe the relations associated to $\boldsymbol{e}_{S'}$ and it is unaware on other relations. In the case when agent i has the tendency to be in a hyponetwork with structure H, it is more likely to observe the structure H after agent i updates e_{ij} whenever the updating process allows the agent to be in a hyponetwork with structure H. The tendency of the agent to prefer to be in H is modelled by introducing a function $Q_{S'}(\bar{\boldsymbol{e}}_{S'})$ such that it takes the value $c \in [0, 1] \setminus \frac{1}{2}$ when $\bar{\boldsymbol{e}}_{S'} \sim H$ and $1 - c$ otherwise. In the case when the value for the variables $\boldsymbol{e}_{S' \setminus \{ij\}}$ are fixed and any value of e_{ij} cannot create a hyponetwork with structure H, the function $Q_{S'}$ is constant on e_{ij}, which implies that the tendency does not make less or more likely to create the relation between i and j. When $c > \frac{1}{2}$, agent i prefers to be in a hyponetwork with structure H and when $c < \frac{1}{2}$ agents prefers not to be in a hyponetwork with structure H.

As an illustration, the function $Q(e_{ij})$ may model the decision of agent i to update the relation e_{ij} when it does not consider the decision of agent j (or any other decision) while the function $Q(e_{ij}, e_{ji})$ may model the decision of agent i while taking into consideration the decision made by agent j regarding the relation e_{ji} but without considering other information. The last function models the tendency of agents to reciprocate a relation while the first one models the tendency of agents to create relations.

In the following, $\{H_1, ..., H_p\}$ denotes p different structures and for each structure we introduce the set $\mathcal{S}_k = \{\boldsymbol{e}_{S'} \subseteq \boldsymbol{e}_S :$ there exists a realisation such that $\bar{\boldsymbol{e}}_{S'} \sim H_k\}$. For each $\mathcal{S}_k$, we define the class of functions $\mathcal{C}_k = \{Q_{S'}^{(k)}\}_{\boldsymbol{e}_{S'} \in \mathcal{S}_k}$ representing the tendency of the agents to belong to hyponetworks with structure H_k. For any $Q_{S'}^{(k)}$, we have that $Q_{S'}^{(k)}(\boldsymbol{e}_{S'}) = c_k$ if $\boldsymbol{e}_{S'} \sim H_k$ and $1 - c_k$ otherwise. We call a class $\mathcal{C}_k$ a simple social mechanism. Although, we will not handle the case of nodes attributes or communities, the above concept can easily be extended to cover these cases (Camacho Guardian, 2016). The above construction covers directed networks, but the undirected case follows straightforward.

2.3 Microscopic foundation

2.3.1 Assumptions at the agent-agent interaction level

The formulation of random graph models at the agent-agent interaction level can be constructed by assuming that the updating process of a relation conditioned on the present and absent relations between the agents can be factorised by some social mechanisms.

Assumption 1. *For all variables $e_{ij} \in \boldsymbol{e}_S$*

$$P(e_{ij} \mid \boldsymbol{e}_{S\setminus ij}) = c_{ij} \prod_{k=1}^{p} \prod_{\substack{e_{ij} \subseteq \boldsymbol{e}_{S'} \\ \boldsymbol{e}_{S'} \in \mathcal{S}_k}} Q_{S'}^{(k)}(\boldsymbol{e}_{S'}) \tag{1}$$

with c_{ij} a constant.

The assumption states that the updating process of any relational variable is influenced by p different simple social mechanisms and that the updating of a relation by an agent is broken in independent decisions occurring in each function Q.

It was shown that the previous microscopic assumption plus some other technical assumptions define an exponential random graph model (Camacho Guardian, 2016), i.e. a probability space $(\Omega_n, \mathcal{G}^n, P_{\boldsymbol{\theta}})$, where Ω_n is all the possible networks with n agents, $\mathcal{G}^n$ is the potential set of Ω_n, and $P_{\boldsymbol{\theta}}$ equals

$$P_{\boldsymbol{\theta}}(\boldsymbol{e}) = \frac{1}{c(\boldsymbol{\theta})} \exp\left(\sum_{k=1}^{p} \theta_k \Gamma_k(\boldsymbol{e})\right). \tag{2}$$

with $\theta_k = \log(\frac{c_k}{1-c_k})$, Γ_k equals the total number of hyponetworks, $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ and $c(\boldsymbol{\theta})$ is a normalising constant. Γ_k is called a network statistics.

Examples of network statistics, for undirected networks, are the number of relations in a network, the number of two-stars and the number of triangles, see Figure 1 (a)-(c). Models including the network statistics number of relations and two-stars are called two-stars models while models including the number of relations and the number of triangles are called triangle models.

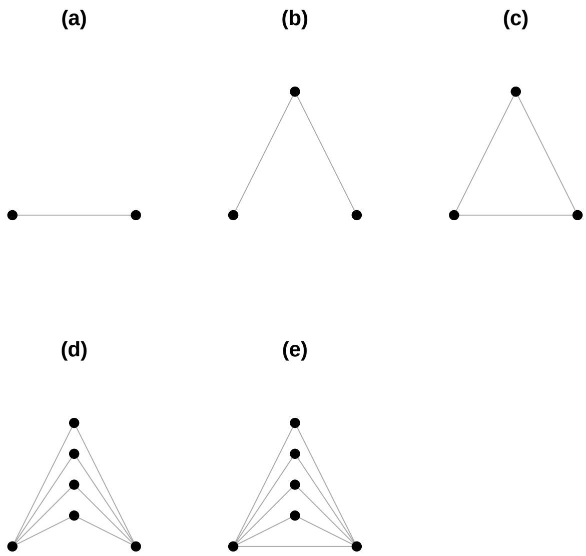


Figure 1: (a)-(c) edge, two star and triangle. (d)-(e) 4-two stars and 4-triangles.

2.3.2 The interdependency between social mechanisms

Since the updating of a relation simultaneously affects the formation of different hyponetworks, some caution is necessary for the assumption that separates the updating process of a relation in independent decisions taking place inside the functions Q. As an example, let us consider the two-star model and let us look at the log of the probability to add a relation.

$$\log(P(e_{ij} = 1 \mid \boldsymbol{e}_{S\setminus ij})) \propto Q^{(1)}(e_{ij}) + \sum_{k \notin \{i,j\}} Q^{(2)}(e_{ij}, e_{i,k}) \tag{3}$$

When the degree of agent i is zero, all the functions $Q^{(2)}$ are inactive, and it is possible to change the log of the probability to add a relation by changing the value of $Q^{(1)}$ while holding constant all other variables. However, when agent i is highly connected, then a large number of functions $Q^{(2)}$ are active, and it is not possible to change the log of the probability to add a relation by changing the value of $Q^{(1)}$ while holding constant all other variables, as it would be expected in regression models. The tendency of agents to be in two stars is magnified when the network is dense, and for large c_2, the updating of a variable will most likely add a relation. On the other hand, the tendency to be in two-stars is absent in the null network, and only $Q^{(1)}$ plays a role in the updating process. As a result, these two mechanisms when having opposing forces can generate distributions that put most of the probability mass on the null- and complete network but a negligible probability mass on the expected value.

The interdependence between some functions Q and observed properties of the two-star model have motivated the development of new network models (Snijders et al., 2006), as we briefly discuss. The problem that some specifications of the two-star models and triangle models place too much probability on a few network is a well-documented problem (Snijders et al., 2006; Li, 2015), and it is called the near-degeneracy problem. The need to solve the near-degeneracy problem motivated the introduction of network statistics such as the number of k-two stars and the number of k-triangles, see Figure 1 (c)-(d). However, due to the large number of parameters needed if introducing all the possible k-two stars (k-triangles) and the observed relationships between the values of the different parameters; these two network statistics motivated more complex network statistics, such as the geometrically weighted edgewise shared partner (GWESP), geometrically weighted dyadic shared partner (GWDSP) and the geometrically weighted degree (GWD) (Hunter, 2007).

Further, the relationships between the different classes of functions Q have found applications to test for constant reciprocity and transitivity on the number of agents in the network (Camacho Guardian A. and Algesheimer, 2017).

The documented relationships between the different classes of functions Q suggest that they cannot be treated isolated, but they are interdependent. Motivated by the interdependency between the social mechanisms, we replace the classical product "$\cdot$" in the conditional distributions (36) by the q-product "$\otimes_q$". This new q-product can be interpreted as interaction terms between the social mechanisms or as a new source of random noise in the model, as we will explain.

2.4 Model

In this section, we present the concept of q-conditional independence and q-Markov random fields. Our definition of q-conditional independence goes in line with the q-independence concept proposed by Sears and Sunehag (2007). The form of the joint distribution of a Markov random field ($q = 1$) is provided by the Hammersley-Clifford (H-C) theorem (Besag, 1974). The application of the Hammersley-Clifford theorem to construct random network models was first proposed by Frank and Strauss (1986) (known as Markov random graph model), and since then it has been applied for constructing larger classes of network models (Pattison and Robins, 2002; Camacho Guardian, 2016). However, the class of network models constructed with Markov random fields are limited to have joint distribution with exponential form; and thus they have a limitation on the probability mass they put on the tails, as we will discuss in the next sections. With the concept of q-conditional independence and q-Markov random fields, we construct the power law random graph model. A generalisation of the Hammersley-Clifford theorem for q-independence was proposed by Sears and Sunehag (2007). A similar notion of q-independence was introduced by Umarov et al. (2008) to establish similar results to the central limit theorem for q-independent random variables but converging to q-Gaussian distributions (Umarov et al., 2008; Umarov and Tsallis, 2016).

2.5 q-Independence and q-Markov random fields

Our proposed generalisation of the microscopic Assumption 1 requires extending the concept of conditional independence. Recall that two set of random variables $\boldsymbol{e}_A$ and $\boldsymbol{e}_B$ are conditional independent on the set of random variables $\boldsymbol{e}_C$ if

$$p(\boldsymbol{e}_A, \boldsymbol{e}_B | \boldsymbol{e}_C) = p(\boldsymbol{e}_A | \boldsymbol{e}_C) \cdot p(\boldsymbol{e}_B | \boldsymbol{e}_C)$$

We do this by substituing the classical product "$\cdot$" in the formula of conditional independence by the q-product. The q-product was introduced by Nivanen et al. (2003); Borges (2004); and it is defined as

$$x \otimes_q y = (0, x^{1-q} + y^{1-q} - 1)_+^{\frac{1}{1-q}}$$

with $(a)_+ = \max(0, a)$.

Definition 1. *We say that two sets of variables $\mathbf{e}_A$ and $\mathbf{e}_B$ are* ***conditional q-independent*** *on the set of variables $\mathbf{e}_C$, denoted as $\mathbf{e}_A \perp\!\!\!\perp_q \boldsymbol{e}_B | \mathbf{e}_C$, if and only if*

$$p(\boldsymbol{e}_A, \boldsymbol{e}_B | \boldsymbol{e}_C) = f(\boldsymbol{e}_A | \boldsymbol{e}_C) \otimes_q f(\boldsymbol{e}_B | \boldsymbol{e}_C).$$

If the set of variables $\boldsymbol{e}_C$ is the empty set, we say that $\boldsymbol{e}_A$ and $\boldsymbol{e}_B$ are q-independent. When $q = 1$, q-conditional independence is reduced to the well-known conditional independence concept.

Often two sets of variables $\boldsymbol{e}_A$, $\boldsymbol{e}_B$ are independent when another set of variables $\boldsymbol{e}_C$ takes a particular value $\bar{\boldsymbol{e}}_C$ but this is not necessarily the case for another realisation of $\boldsymbol{e}_C$. This

The importance of q-Gaussians stems from the fact that they are a generalisation of the Gaussian distribution but favouring heavier tails.

type of independence was introduced in the network literature by the term *partial conditional independence* (Pattison and Robins, 2002), and it has found applications in the construction of some network statistics (Snijders et al., 2006), e.g. GWESP, GWDSP.

Definition 2. *We say that two sets of variables $\boldsymbol{e}_A$ and $\boldsymbol{e}_B$ are* ***partial q-independent*** *on $\boldsymbol{e}_C = \bar{\boldsymbol{e}}_C$, denoted as $\boldsymbol{e}_A \perp\!\!\!\perp_q \boldsymbol{e}_B | \boldsymbol{e}_C = \bar{\boldsymbol{e}}_C$, if and only if*

$$p(\boldsymbol{e}_A, \boldsymbol{e}_B | \boldsymbol{e}_C = \bar{\boldsymbol{e}}_C) = f(\boldsymbol{e}_A | \boldsymbol{e}_{\boldsymbol{C}} = \bar{\boldsymbol{e}}_C) \otimes_q f(\boldsymbol{e}_B | \boldsymbol{e}_C = \bar{\boldsymbol{e}}_C).$$

The dependency structure between a set of random variables $\boldsymbol{e}_S$ can be represented by an undirected graph $G(V, E)$ (Kevin, 2012). Let $G(V, E)$ represents a generic graph with the set of vertices equals the set $\boldsymbol{e}_S$ but there is no restriction on the edges E. The undirected graph encodes (a subset of) conditional independence relations between the variables via the following Markov properties.

Definition 3. *We say that an undirected graph G satisfies the q*-**pairwise Markov property** *if and only if for any pair $e_{ij}, e_{i'j'}$ of non-adjacent nodes, we have that*

$$e_{ij} \perp\!\!\!\perp_q e_{i'j'} | \boldsymbol{e}_{S \setminus \{ij, i'j'\}}$$

The neighbourhood for a node e_{ij} in the graph G is defined as all the nodes adjacent to e_{ij} and it is denoted by $N(ij)$.

Definition 4. *We say that an undirected graph G satisfies the q*-**local Markov property** *if and only if for each e_{ij} node, we have that*

$$e_{ij} \perp\!\!\!\perp_q \boldsymbol{e}_{S \setminus N(ij)} | \boldsymbol{e}_{S \setminus N(ij) \setminus \{ij\}}$$

Definition 5. *We say that the q-dependency graph G satisfies the q*-**Global Markov property** *if and only if for each pair of disjoint subsets $\boldsymbol{e}_A, \boldsymbol{e}_B$ such that $\boldsymbol{e}_C$ separates $\boldsymbol{e}_A$ and $\boldsymbol{e}_B$, we have* *that*

$$\boldsymbol{e}_A \perp\!\!\!\perp_q \boldsymbol{e}_B | \boldsymbol{e}_C$$

Next, we extend the definition of Markov random field to cover the concept of q-conditional independence as follows.

Definition 6. *A* ***q-Markov random field*** *is a set of random variables $\boldsymbol{e}_S$ with and undirected graph $G(V, E)$ satisfying the q-local Markov property.*

2.6 Power law random graph models (PRGM)

In this section, we present the first construction of the power law random graph model by introducing assumptions at the agent-agent interaction level.

Assumption 2. *For any variable $e_{ij} \in \boldsymbol{e}_S$, we have that the conditional probability on the remaining variables equals*

$$P(e_{ij} \mid \boldsymbol{e}_{S\setminus ij}) = c_{ij}\Big(\bigotimes_{k=1}^{p}{}_q \bigotimes_{\substack{e_{ij}\subseteq \boldsymbol{e}_{S'} \\ \boldsymbol{e}_{S'}\in \mathcal{S}_k}} {}_q Q^{(k)}_{S'}(\boldsymbol{e}_{S'})\Big) \tag{4}$$

with c_{ij} a constant.

As in Assumption 1, the updating process of a relational variable is factorised by p simple social mechanisms. The q-product adds another source of correlation between the variables, but it does not introduce any new social mechanisms to the model. Motivated by our previous observation that the values of the functions Q are highly interdependent, e.g. the formation of a triangle causes the formation of a 2-star and a relation. We interpret the q-product as the introduction of interaction terms between the functions Q.

To understand the sources of the interaction terms appearing with q, we can look at the log

The conditional probability in (4) can be written in terms solely of the q-product, see Supplementary Materials.

of the conditional probability of updating a relation.

$$\log(P(e_{ij} \mid \boldsymbol{e}_{S\setminus ij})) \propto \sum_{k=1}^{p} \sum_{\substack{e_{ij}\subseteq \boldsymbol{e}_{S'} \\ \boldsymbol{e}_{S'}\in \mathcal{S}_k}} \log_q(Q_{S'}^{(k)}(\boldsymbol{e}_{S'})) - \frac{(1-q)}{2}\Big(\sum_{k=1}^{p} \sum_{\substack{e_{ij}\subseteq \boldsymbol{e}_{S'} \\ \boldsymbol{e}_{S'}\in \mathcal{S}_k}} \log_q(Q_{S'}^{(k)}(\boldsymbol{e}_{S'}))\Big)^2 + \mathcal{O}((1-q)^2) \quad \text{as } q \to \tag{5}$$

where

$$\log_q(x) = \begin{cases} \log(x) & q = 1 \\ \frac{x^{1-q}-1}{1-q} & q \neq 1 \end{cases}$$

$\log_q(x)$ is known as the q-logarithm function, and it equals the logarithm when $q = 1$.

From the previous equation, we observe that when $q = 1$, the log odds of the probability is broken in independent tendencies of agents to belong to particular hyponetworks. However, when $q \neq 1$, interaction terms between the tendencies is added to the model, with the influence of the interaction term being of order $\mathcal{O}(1 - q)$. When q increases, the influence the interaction terms have on the updating of relation increases; and the influence the social mechanisms have in the process diminish. As a result, the log odds of the probability of adding a relation converges to one as q tends to infinity (see Supplementary Materials).

Another interpretation of the q-product is an idiosyncratic correlation between the variables that is not caused by any particular preference of the agents to belong to hyponetworks with specific structures. In this context, the elements having $(1 - q)$ on the right side of Equation (5) are similar to a random noise unobserved by the researcher, but at the level of the social mechanisms themselves.

As previously mentioned, the description at the agent-agent interaction level is done with the q-algebra while the distribution over the networks is with the generalised q-exponential and q-logarithm functions. The next result shows that the form of the probability distributions over the networks arising from Assumption 2 are power laws, with the exponential random graph model as a special case when $q = 1$.

Result 1. *If Assumption 2 is satisfied, then the probability distribution over the networks takes the form*

$$P(\boldsymbol{e}) = \frac{1}{c(\boldsymbol{\theta}, q)} \exp_q \Big(\sum_{k=1}^{p} \theta_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta}) \Big) \tag{6}$$

with $\theta_k = -2h_q(1 - c_k)$, $\varphi(\boldsymbol{\theta}) = \sum_{k=1}^{p} N_k \log_q(h_q^{-1}(-\boldsymbol{\theta}_k))$ *and* $h_q(x) = \frac{1}{2}(\log_q(1+x) - \log_q(1-x))$,

$$\exp_q(x) = \begin{cases} \exp(x) & q = 1 \\ \max(0, (1 + (1-q)x))^{\frac{1}{1-q}} & q \neq 1 \end{cases}$$

$\exp_q(x)$ *is known as the* q*-exponential function, and it equals the exponential function when* $q = 1$.

Random graph models are often constructed by describing the dependency structure between the relational variables using a Markov random field model. For this, let us define the dependency graph of a PRGM satisfying Assumption 2 as follows.

Definition 7. *For a PRGM defined by the* q*-product, we define its* q*-dependency graph* G_q *as follows: 1. each node represents a variable* e_{ij}*. 2. two nodes* e_{ij} *and* $e_{i'j'}$ *are connected in the dependency graph if and only if there exists* k *and* $\boldsymbol{e}_{S'} \in \mathcal{S}_k$ *such that* $e_{ij}, e_{i'j'} \in \boldsymbol{e}_{S'}$.

Result 2. *For any PRGM, its dependency graph* G_q *satisfies the pairwise and local* q*-Markov property. When* $q = 1$*, the global property also holds.*

2.6.1 q-Bernoulli random graph model

When the q-dependency graph is the null graph, we call the power law random graph model simply a q-Bernoulli random graph model. The class of q-Bernoulli random graph models is defined by a set of n agents, a parameter c defining the tendency of agents to relate with others

θ_k is also equal to $\log_q(c \ominus_q (1-c))$ where $\ominus_q$ is the q-substraction operation defined in Borges (2004).
A weaker version of the global property is satisfied by the PRGM as it shown is the Supplementary Materials.

and the parameter q, $B(n, q, c)$. Thus, the only social mechanism taking place is the tendency of agents to relate with other agents, and the relational variables are correlated solely by the idiosyncratic correlation q (or the interaction terms between the tendencies of agents to create relations). For $q = 1$, this class of models tends to place too much probability mass on networks having a number of relations close to the expected value and not so much probability on the tails of the distribution, as we discussed next. In particular, fitting a model to a set of networks using MLE, and then simulating networks from the fitted model will most likely generate networks with a number of relations close to the average number of relations. However, if there is too much variability in the number of relations in the observed networks, some networks will have a negligible probability of occurring according to the fitted model.

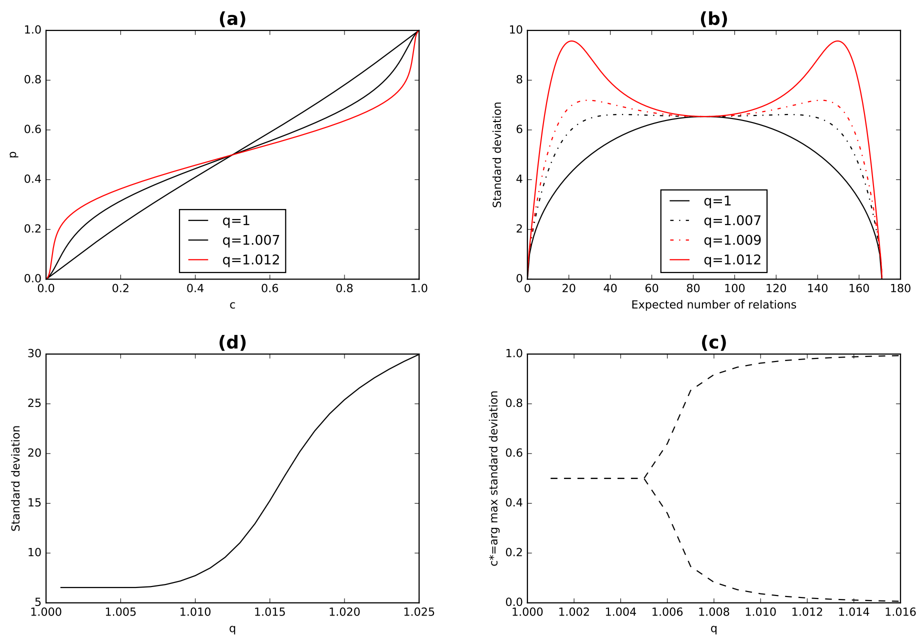


Figure 2: Plot (a) shows the relationship between the marginal probability that there is a relation between two agents against the parameter c. Plot (b) shows the standard deviation against the expected number of relations for the models $B(19, q, c)$. Plot (c) shows the maximum standard deviation for the class $B(19, q, c)$ as a function of q. Plot (d) shows the arguments c that maximize the standard deviation for the classes $B(19, q, c)$.

In Figure 2 (a), we plot the probability of observing a relation between two agents, $p_c = P_c(e_{ij} = 1)$, against the parameter c. For $q = 1$, we observe a linear relationship between p_c

and c, but this does not longer hold for $q > 1$ due to the dependencies between the relational variables. $p_c(q)$ is an increasing function on q in the interval $c \in (0, \frac{1}{2})$, a decreasing function in the interval $(\frac{1}{2}, 1)$, and it is a constant function when c equals $0, \frac{1}{2}$ or 1. This is consistent with our previous observation that the odds of the probability of creating a relation converges to 1 as q tends to infinity. In Figure 2 (b), we observe how by increasing q, it is possible to increase the standard deviation on the number of relations. For $q = 1$, the standard deviation of the 1-Bernoulli random graph model equals $\sqrt{n\, c\,(1-c)}$ (n equals the number of agents in the network), which is bounded by above by $\sqrt{n/4}$. In Figure 2 (c), we show the maximum standard deviation for the class of q-Bernoulli random graph models and different values of q. The model maximising the standard deviation is unimodal up to a certain $q^{\star}$, and afterwards, the probability of the null network becomes non- negligible (see supplementary materials). For large q, the large standard deviation is achieved with models that are a combination of a deterministic model that place too much probability mass on the null network (or the complete network) and a Bernoulli random graph model with $p = \frac{1}{2}$.

The upper bound on the standard deviation for the Bernoulli random graph model can be understood as the problem of placing too much probability mass on a few class of networks, and by increasing q it is possible to increase the variability without introducing new social mechanisms to the model. However, for large q and small c, we observe again the problem of placing to much probability on a few classes of models, but as a result of the strong correlations, the distribution becomes bimodal.

The value of c maximizing the standard deviation equals $\frac{1}{2}$ when $q = 1$, and there are two c maximizing the standard deviation for $q > 1$: one ($c^m(\bar{q})$) smaller than $1/2$, and the other ($c^M(q)$) larger than $1/2$, see Figure 2 (d). The limit of $c^m(q)$ when q tends to infinity is zero, while the limit of $c^M(q)$ equals 1.

Lastly, let us consider the $\tilde{c}(q)$ that defines the q-Bernoulli random graph model with expected number of relations equals a constant $const$. $\tilde{c}(q)$ is a decreasing function of q for $\frac{const}{n}$ in the interval $(0, \frac{1}{2})$, see Figure 3 (b), and a decreasing function in the interval $(\frac{1}{2}, 1)$. $\tilde{c}(q)$ is a constant function of q when $\frac{const}{n}$ equals $0, \frac{1}{2}$ or 1.

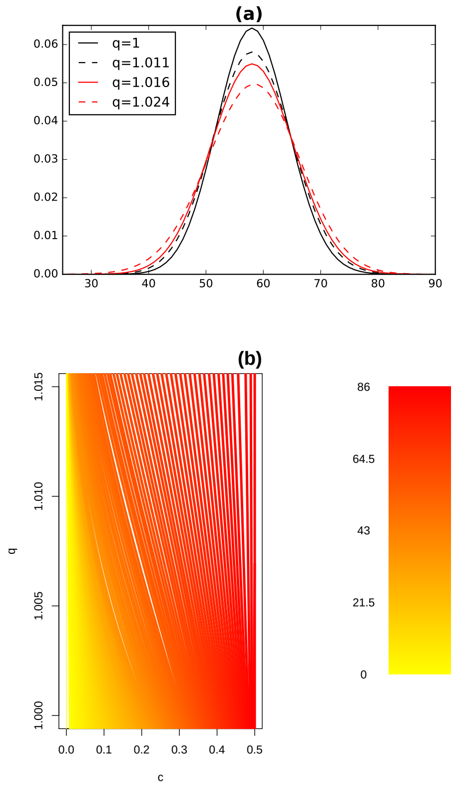


Figure 3: Plot (a) shows the distributions of different q-Bernoulli random graph model with the expected number of relations fixed to 58. A curve in Plot (b) shows the parameters c and q where the expected value of the number of relations for the q-Bernoulli random graph model is constant. The red lines is when the expected value is close to 85.5, and the yellow curves when the expected value is close to zero.

2.6.2 q-Markov graph model

The class of PRGM which have as a constraint that if $e_{ij}, e_{i'j'} \subseteq \boldsymbol{e}_{S'} \in \mathcal{S}_k$, then $\{i, j\} \cap \{i', j'\} \neq \emptyset$ are called q-Markov graph model. A q-Markov graph model only models q-dependencies between relational variables that are adjacent and they are a generalisation of the models proposed by Frank and Strauss (1986). For the undirected networks, special cases of q-Markov graph model are the q-two star- q-triangle models. As previously discussed, for the 1-triangle (1-two star) model, it has been observed that these models often places too much probability on networks with low density and on networks with a density close to 1, and they are associated with the near-degeneracy problem. The reason is that these models consist of two competitive mechanisms, one that controls for the tendency of agents to be connected and the other of having relations that are part of two-stars, or triangles. For the chosen parameters, in the below three models, the first mechanism makes the probability of adding a relation quite unlikely, while the second mechanism makes more likely to add a relation when adding the relation adds a two-star/triangle to the network. In the extreme cases when $q = 1$ and the network is the complete network, it is quite unlikely that the updating of a relational variable culminates on the removal of a relation, and in the null network it is quite unlikely that the updating of variables culminate in the formation of new relations. Simulating networks using Monte Carlo Markov chain methods will independent of the initial condition almost certainly move the network to one of these two extreme networks, and it will stay there for a long period of time (Snijders et al., 2006). For $q > 1$, the idiosyncratic correlation shifts the odds closer to 1, and it decreases the effect of the competitive mechanisms. As a result, the social mechanisms do not necessarily move the network to the two extreme cases, but it certainly affects the distribution.

As a first example, we construct a 1-two star model with five agents and parameter c_1 close to zero and c_2 larger than $\frac{1}{2}$, where c_1, c_2 model the formation of relations and two-stars, respectively. The model puts most of its probability mass on the null- and complete networks, while it puts negligible probability mass in values close to the expected value. When q increases, Figure 4 (a), the issue of placing most of the probability mass on a few networks starts to diminish, and at certain point the q-stars model places its probability mass approximately evenly. However, if we continue to increase q, the shape of the distributions goes from a bimodal to unimodal. The unimodal distributions for large q are the result that the effects of the two social mechanisms are overshadowed by the idiosyncratic correlation, which makes the odds of adding a relation equal to 1. As second and third examples, we construct two 1-triangles models with 15 agents and parameter c_1 close to zero and c_2 larger than $\frac{1}{2}$, where c_1, c_2 model the formation of relations and triangles, respectively. Both models are putting negligible probability mass on networks that are neither the null network nor the complete network. When q increases, the probability mass on both models starts to move away from the extreme cases, as shown in Figure 4 (b)-(c). For large q, the distribution becomes unimodal. In general, the models obtained with large q have the unfortunate property of placing most of its probability mass on a few types of networks, e.g. networks which number of relations are approximately equal to the expected number of relations of a Bernoulli random graph model with $c = \frac{1}{2}$. Therefore, near-degenerate models occur when $q = 1$ and for large q.

As previously discussed, bimodal distributions have mostly been documented in models placing most of its probability on a few networks- near degenerate models; and as a consequence, they have been associated with misspecified model. Our previous examples, though, illustrates that this phenomenon does not solely occur in near-degenerate models. For the family of q-Markov graph models, the dependencies between the relational variables give rise to bimodal-, skewed-, and flat distributions, with the shape being the aftermath of the strength of the dependency of the decision made by the agents. The strength is often the problem to be studied, and it is not controlled by researchers, nor strong assumptions can be imposed on it. When the strength of the dependencies undermines Gaussian approximations of the network statistics, it also invalidates the construction of p-values and confidence intervals. Also, the non-concentration of the probability around the expected value highlights the difficulty to justify the estimation of models assuming that the observed statistics are approximately equal to the expected value under the true model.

2.7 Maximum Entropy Principle

Park and Newman proposed a derivation of the ERGM based on the concept of entropy (Park and Newman, 2004), $H(\boldsymbol{e}) = \sum_{\tilde{\boldsymbol{e}} \in \mathcal{G}^n} P(\tilde{\boldsymbol{e}}) \log \big(P(\tilde{\boldsymbol{e}})\big)$. Entropy connects the microscopic level, at which agents interact, and the macroscopic level, on which laws describing the behaviour of the system are formulated. It was first proposed in statistical physics by Boltzmann, and it was formulated as a measure of uncertainty and axiomatised by Shannon (1949) and

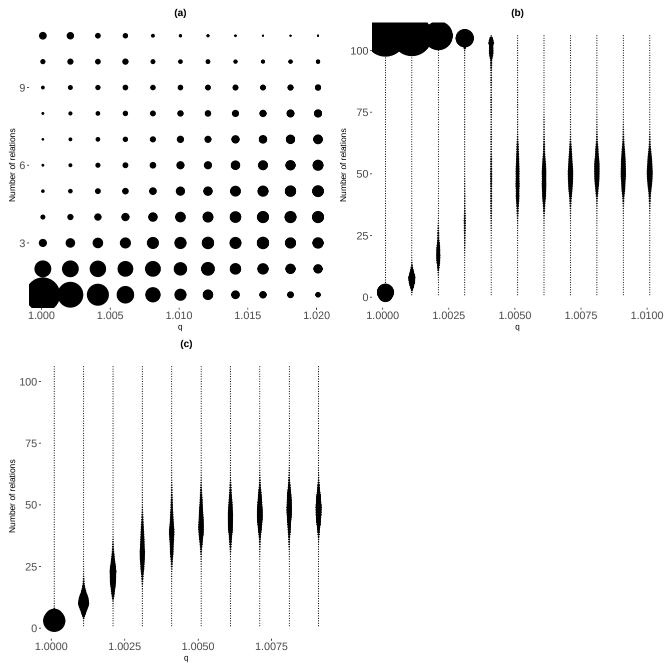


Figure 4: Figure (a) shows the probability distribution of the 1-two star model with five agents and parameters $(c_1, c_2) = (0.047, 0.72)$. The distribution was computed by simulating $100,000$ networks using a Monte Carlo Markov Chain algorithm. Figures (b), (c) shows triangle models with 15 agents and parameters $(c_1, c_2) = (9e-4, 0.73), (0.02, 0.72)$ respectively. Both distributions were computed by simulating $250,000$ networks using a Monte Carlo Markov Chain algorithm. For the three points, the size of the points is proportional to the probability of observing a network with the given number of relations.

Aleksandr and Khinchin (1949). From the assumptions to justify the Boltzmann-Shannon entropy, the additive assumption is the least intuitive assumption, and its relaxation/modification established new classes of entropies such as the Renyi- and Tsallis entropy (Rényi et al., 1961; Tsallis, 1988). Additivity states that the entropy of a system consisting of the combined relational variables $(\boldsymbol{e}_{S'}, \boldsymbol{e}_{S''})$ equals the entropy of the system consisting only of the relational variables $\boldsymbol{e}_{S'}$ plus the average of entropy of the variables $\boldsymbol{e}_{S''}$ given the variables $\boldsymbol{e}_{S'}$: $H(\boldsymbol{e}_{S'}, \boldsymbol{e}_{S''}) = H(\boldsymbol{e}_{S'}) + H(\boldsymbol{e}_{S''}|\boldsymbol{e}_{S'})$.

However, some caution is necessary on the additivity for finite systems or systems exhibiting long-range correlations (Tsallis, 2009). Following Boltzmann approach, entropy is additive in the thermodynamic limits and using Stirling approximations (Kakorin, 2009; Riek and Sobol, 2016). In large deviation theory, entropy is called the rate function; and it controls the rate of the exponential decay of rare events, and with the exponential decay being a consequence of the additive property of the Boltzmann-Shannon entropy (Touchette, 2009).

Thus, we conjecture that if the goal of a class of random network models is to decrease the concentration of probability mass around its mode(s) without increasing the number of social mechanisms in the model, network models must go beyond formulations based on Boltzmann-Shannon entropy. The main result of this section shows that PRGM finds another theoretical foundation in Tsallis entropy. Tsallis is a generalisation of the Boltzmann-Shannon entropy (Tsallis, 1988), and it has been applied to a variety of complex systems, such as signal analysis (Chen and Li, 2014), ecology (Komori and Eguchi, 2015), train delays (Briggs and Beck, 2007), the stock market (Iliopoulos et al., 2015) just to mention some.

Result 3. *Let*

$$H_q(\boldsymbol{e}) = \sum_{\boldsymbol{e}\in\mathcal{G}^n} P(\boldsymbol{e}) \log_q \left(\frac{1}{P(\boldsymbol{e})}\right) \tag{7}$$

Subject to

$$\sum_{\boldsymbol{e}\in\mathcal{G}^n} P(\boldsymbol{e}) = 1 \tag{8}$$

and

$$\sum_{\boldsymbol{e}\in\mathcal{G}^n} \Gamma_k(\boldsymbol{e}) P_q(\boldsymbol{e}) = \bar{\Gamma}_k \tag{9}$$

for some constants $\bar{\Gamma}_k$ *and*

$$P_q(\boldsymbol{e}) = \frac{P^q(\boldsymbol{e})}{\sum_{\boldsymbol{e}\in\mathcal{G}^n} P^q(\boldsymbol{e})}$$

If the solution to the Maximization problem exists, it is of the form:

$$P(\boldsymbol{e}) = \frac{1}{c(\boldsymbol{\theta}, q)}\Big(1+(1-q)\big(\varphi(\boldsymbol{\theta})-\sum_{k=1}^{p}\theta_k\Gamma_k(\boldsymbol{e})\big)\Big)^{\frac{1}{1-q}} = \frac{1}{c(\boldsymbol{\theta}, q)}\exp_q\Big(\sum_{k=1}^{p}\theta_k\Gamma_k(\boldsymbol{e})+\varphi(\boldsymbol{\theta})\Big) \tag{10}$$

for some $\boldsymbol{\theta} = (\theta_1, ..., \theta_p) \in \mathbb{R}^p$, $\varphi(\boldsymbol{\theta}) =< \boldsymbol{\theta}, \bar{\boldsymbol{\Gamma}} >$ *and* $c(\boldsymbol{\theta}, q)$ *a normalising constant.*

The probability P_q is called the q-escort probability distribution. As discussed by Tsallis et al. (2009), when dealing with distributions that decay slowly (e.g. power laws), it is not possible to characterise them with standard moments (expectation and variance).

2.8 Empirical Results

In this section, we estimate q-Bernoulli random graph models. Our data consists of 57 networks with each network representing friendship relations between high school students in a classroom in Switzerland (Algesheimer, 2016). We use the first wave of the data, and it comprises students in the 7th and 9th grades. We also include 36 networks with each network representing friendships relations between high school students in a classroom in two schools in the US (McFarland, 2001). We also use the first wave, and the classrooms are from 10th and 11th grade. We combine the two datasets, and then we group the networks according to the number of students. The reason to combine both datasets are twofold: (i) each network is treated as a single observation and (ii) network models are defined on the set of networks with a specific number of agents in the network. As the number of networks in each group does not go above 11 networks, our empirical analysis is meant to be exploratory in nature. The size of the network ranges from 12 to 30 students, and the total number of students in Swiss classrooms is 1160 while the total number of students in US classrooms is 790. We estimate two models to each group of networks using MLE. The first estimated model is a simple 1-Bernoulli random graph model, and the second estimated model is a q-Bernoulli random graph model with q a parameter to be estimated.

The estimated q-Bernoulli random graph model for one observation is always when $q = 1$ and this is the case for the group of networks with sizes $14, 29$ and 30. For the groups of networks with not so much variability on the observed number of relations and the observed number of relations are close to the average, the parameter q does not seem to improve the fitting power of the data. This is not a surprise as Bernoulli random graph models tend to put most of their probability mass around the expected value. However, this goodness-of-fit on the observed number of relations should be treated with caution, as it may be a consequence of overfitting.

q-Bernoulli random graph models outperform 1-Bernoulli random graph models when there is variability in the total number of relations in the classroom. For the groups of networks with 12, 16, 19-28 students, the estimated values of q range between 1.00675 and 1.077. The estimated q are significant at the 0.05-level for the group of networks of sizes 12, 20-24, 26-28, see Figure 5 (a). However, the two models estimated with the set of networks of size 18, 22 posse the undesirable property of placing a non-negligible amount of probability mass on the null network (0.007, 0.012).

The correlation between the estimated q and the size of the networks is -0.9. The negative correlation is explained by the fact that as the number of agents increases and c is constant, the influence q has on the model increases, i.e. the model gets closer to $B(n, q, \frac{1}{2})$, and it is necessary to adjust q to keep the expected number of relations constant. Contrary to the observation that the parameter c is negatively correlated with n for the $B(n, 1, c)$ Camacho Guardian (2016), the estimated parameter c is positively correlated with the size of the network, 0.68. The change in sign of the correlation is partially understood by the fact that the marginal probability of having a relation between two agents decreases as the number of agents increases, and it is the responsible for the negative correlation between c and n. For $B(n, q, c)$, the marginal probability is controlled by the parameter c and q, and the marginal probability of having a relation between two agents is negatively correlated with c, -0.63, and positively correlated with q, 0.79.

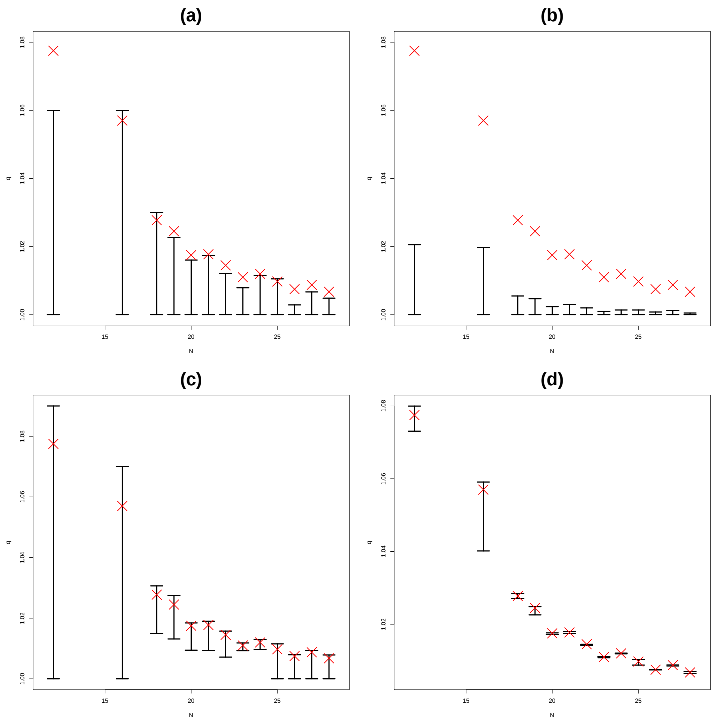


Figure 5: The bars in Figures (a) and (c) show the 95% confidence interval under the null hypothesis that q equals 1 and the sample size equals $5, 1000$, respectively. The bars in Figure (b) and (d) show the 95% confidence interval under the null hypothesis that q equals the estimated value and the sample size equals $5, 1000$, respectively. In all Figures, the red cross represents the estimated parameter.

In Figure 6, we see the distribution of the two estimated models for networks with 19, 20, 23 and 26 agents. We observe that the probability mass of the estimated 1-Bernoulli random graph model is always highly concentrated around the expected value, but this concentration of the probability mass around its expected value is not longer observed for the estimated q-Bernoulli random graph model. In Figure 7, we see that the probability of observing networks with the number of relations equal or smaller than the number of relations of the least dense network in the group is always underestimated for each estimated 1-Bernoulli random graph model. The same holds for the probability of observing networks with the number of relations equal or larger than the number of relations of the densest observed network. The fact that the variance for the q-Bernoulli random graph models is in many cases significantly higher than for the 1-Bernoulli random graph models is consistent with our previous observation that network models based on the classical entropy place negligible probability mass on their tails.

Although the estimator $\hat{c}$ is consistent when q is fixed, we do not have a priori knowledge about q and we consider q a parameter to be estimated. As a result, there are some issues concerning our estimation procedure. First, the estimated q is by definition greater or equal than 1, and thus the estimator $\hat{q}$ is biased when the true value of q is 1. For the uninteresting case when the sample consists of one network, $\hat{q}$ is deterministic with $\hat{q} = 1$, and thus it is biased when the true value of q is greater than 1 and unbiased when $q = 1$. However, as shown in Figure 5 (a), the estimated values suggest that the true value is different from 1. Further Figures 5 (a)-(b) show that as n grows, the estimated value converges to the true value when the true value is q, while Figure (c)-(d) shows that convergence to the true value also occurs when $q \neq 1$.

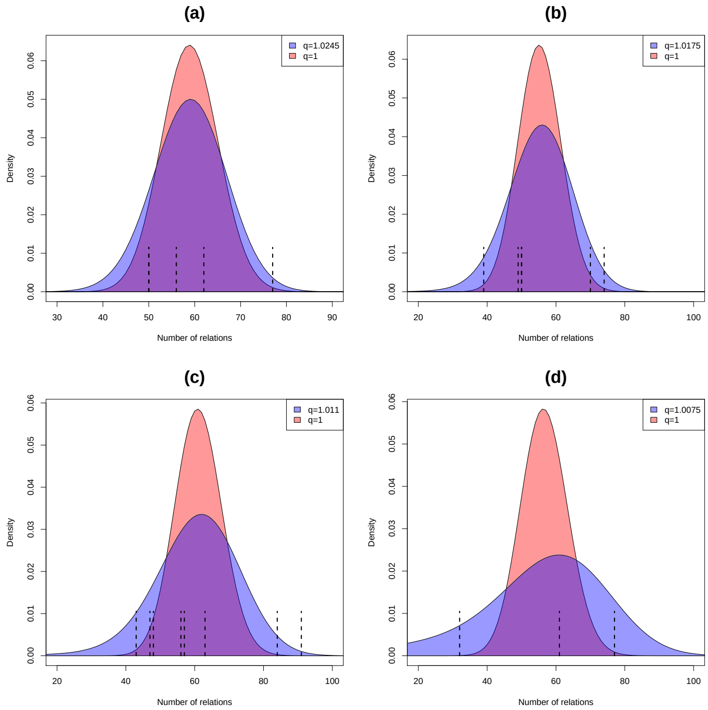
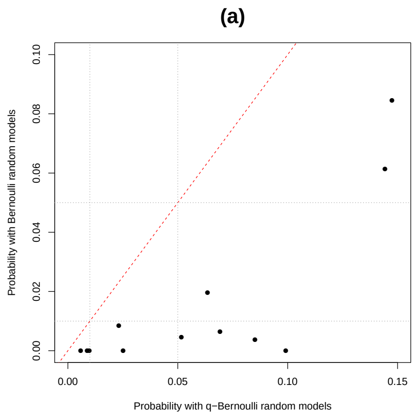


Figure 6: Figures (a),(b), (c) and (d) show the probability distribution of the estimated models for the group of networks with 19, 20, 23 and 26 agents, respectively. The red vertical lines represent observed numbers of friendships in different classrooms. In all cases, we have that the q-Bernoulli random graph model decreases the probability mass around the expected value.



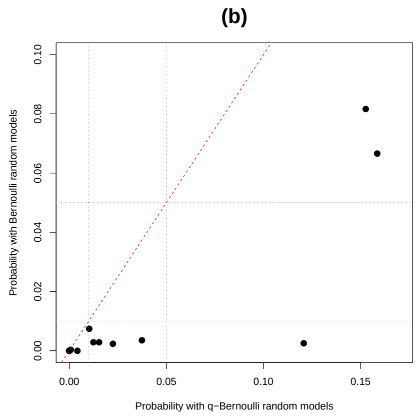


Figure 7: Figure (a) shows the probability of observing a network with less or equal number of friendships than the least dense observed network. Figure (b) shows the probability of observing a network with more or equal the number of friendships than densest observed network. The numbers next to the points represents the size of the networks used to estimate the models.

2.9 Conclusions

The novelty of PRGM is that it extends existing network models by introducing interdependencies between social mechanisms underlying the formation of a network. We do this with the q-algebra with q a parameter that can be interpreted as interaction terms between the social mechanisms in the model or as a new source of random noise (idiosyncratic correlations). Further, we generalise the concept of Markov random field, and we generalise two well-known classes of network models: Bernoulli random graph models and Markov random graph model. We show that Bernoulli random graph models have the inconvenient property of placing too much probability mass on a few networks. Thus, the problem of network models placing too much probability mass on a few networks not only occurs in models with bimodal distributions having a negligible probability mass on the expected value, as it has been extensively documented (Snijders et al., 2006; Robins et al., 2007; Li, 2015). It also occurs in models placing most of its probability mass around the expected value. We show that in our generalised Bernoulli random graph model, it is possible to increase the probability mass on the tails without adding new social mechanisms to the model but by tuning the parameter q. With the help of the generalised Markov random graph model, we showed existing issues in the construction of factual evidence using network data. The observed non-Gaussian distributions under simple dependency assumptions highlight issues on the underlying assumptions needed to construct p-values and confidence intervals.

Although, Boltzmann-Shannon entropy have provided a solid theoretical framework for the construction of network models. The underlying assumptions of the Boltzmann-Shannon entropy are unlikely to hold in complex systems with a small number of constituents and when there are strong correlations between the actions of the constituents. Therefore, it is necessary to construct network models that are based on generalisation of the Boltzmann-Shannon entropy. PRGM partially address this issue, as we show that they can be constructed using Tsallis entropy. In particular, exponential random graph model are a special case of PRGM when $q = 1$.

Some challenges were not addressed in this paper. First, estimation in PRGM are at least as difficult as ERGM, and our algorithms do not allow us to perform MCMC-MLE to models beyond the q-Bernoulli random graph model. This is important if PRGM is aimed to become a useful tool for statistical analysis of network data and not merely a model for analysing properties of network models. Second, as there is no previous study inquiring the possible values of q for social networks, in our empirical study q is a parameter to be estimated. However, due to the limitation of our datasets and lack of prior knowledge of the possible values, some caution is necessary for interpreting the documented values of q. Third, we believe that alternative mathematical frameworks are plausible and needed to better understand existing methodologies used to interpret network data.

References

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., and Liu, B. (2011). Predicting flu trends using twitter data. In *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, pages 702–707. IEEE.

Aleksandr, I. and Khinchin, A. (1949). *Mathematical foundations of statistical mechanics*. Courier Corporation.

Algesheimer, R., D. E. C. J. K. M. R. H. (2016). Values in schools project.

Amari, S.-i. and Ohara, A. (2011). Geometry of q-exponential family of probability distributions. *Entropy*, 13(6):1170–1185.

Andriani, P. and McKelvey, B. (2007). Beyond gaussian averages: redirecting international business and management research toward extreme events and power laws. *Journal of International Business Studies*, 38(7):1212–1230.

Andriani, P. and McKelvey, B. (2009). Perspective-from gaussian to paretian thinking: causes and implications of power laws in organizations. *Organization Science*, 20(6):1053–1071.

Axelrod, R. (1997). The dissemination of culture a model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226.

Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528. ACM.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144):1236498.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.

Betz, F., Hautsch, N., Peltonen, T. A., and Schienle, M. (2016). Systemic risk spillovers in the european banking and sovereign network. *Journal of Financial Stability*, 25:206–224.

Block, P., Stadtfeld, C., and Snijders, T. A. (2016). Forms of dependence comparing saoms and ergms from basic principles. *Sociological Methods & Research*, page 0049124116672680.

Boda, Z. and Néray, B. (2015). Inter-ethnic friendship and negative ties in secondary school. *Social Networks*, 43:57–72.

Borges, E. P. (2004). A possible deformed algebra and calculus inspired in nonextensive thermostatistics. *Physica A: Statistical Mechanics and its Applications*, 340(1):95–101.

Briggs, K. and Beck, C. (2007). Modelling train delays with q-exponential functions. *Physica A: Statistical Mechanics and its Applications*, 378(2):498–504.

Buchmann, T., Hain, D., Kudic, M., and Müller, M. (2014). Exploring the evolution of innovation networks in science-driven and scale-intensive industries: New evidence from a stochastic actor-based approach. Technical report, IWH Discussion Papers.

Calvó-Armengol, A. and Zenou, Y. (2004). Social networks and crime decisions: The role of social structure in facilitating delinquent behavior. *International Economic Review*, 45(3):939–958.

Camacho Guardian, A. (2016). On evidence and social networks.

Camacho Guardian A., Tessone, C. and Algesheimer, R. (2017). Statistical analyses on small networks.

Centola, D. (2011). An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272.

Chandrasekhar, A. G. and Jackson, M. O. (2015). A network formation model based on subgraphs. *Available at SSRN*.

Chen, J. and Li, G. (2014). Tsallis wavelet entropy and its application in power signal analysis. *Entropy*, 16(6):3009–3025.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

De La Haye, K., Robins, G., Mohr, P., and Wilson, C. (2011). Homophily and contagion as explanations for weight similarities among adolescent friends. *Journal of Adolescent Health*, 49(4):421–427.

Ellis, R. S., Newman, C. M., and Rosen, J. S. (1980). Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 51(2):153–169.

Eubank, S., Guclu, H., Kumar, V. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., and Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.

Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational and Behavioral Statistics*, 12(2):101–128.

Goldenberg, J., Libai, B., and Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3):211–223.

Haldane, A. G. and May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, 469(7330):351–355.

Handcock, M. S., Robins, G., Snijders, T., Moody, J., and Besag, J. (2003). Assessing degeneracy in statistical models of social networks. In *Journal of the American Statistical Association*.

Hayashi, F. (2000). Econometrics.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

Huitsing, G. and Veenstra, R. (2012). Bullying in classrooms: Participant roles from a social network perspective. *Aggressive Behavior*, 38(6):494–509.

Hunter, D. R. (2007). Curved exponential family models for social networks. *Social networks*, 29(2):216–230.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3).

Iliopoulos, A., Pavlos, G., Magafas, L., Karakatsanis, L., Xenakis, M., and Pavlos, E. (2015). Tsallis q-triplet and stock market indices: The cases of s & p 500 and tvix. *Journal of Engineering Science and Technology Review*, 8(1):34–40.

Kakorin, S. (2009). Revision of boltzmann statistics for a finite number of particles. *American Journal of Physics*, 77(1):48–53.

Karwa, V., Petrović, S., and Bajić, D. (2016). Dergms: Degeneracy-restricted exponential random graph models. *arXiv preprint arXiv:1612.03054*.

Kevin, M. (2012). Machine learning: a probabilistic perspective.

Komori, O. and Eguchi, S. (2015). Maximum power entropy method for ecological data analysis. In *BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING (MAXENT 2014)*, volume 1641, pages 337–344. AIP Publishing.

Lewis, K. (2013). The limits of racial prejudice. *Proceedings of the National Academy of Sciences*, 110(47):18814–18819.

Li, K. (2015). *Degeneracy, Duration, and Co-evolution: Extending Exponential Random Graph Models (ERGM) for Social Network Analysis*. PhD thesis.

Lobão, T. C. P., Cardoso, P. G., Pinho, S. T., and Borges, E. P. (2009). Some properties of deformed q-numbers. *Brazilian Journal of Physics*, 39(2A):402–407.

Maslov, S. and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913.

McFarland, D. A. (2001). Student resistance: How the formal and informal organization of classrooms facilitate everyday forms of student defiance. *American Journal of Sociology*, 107(3):612–678.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827.

Moody, J. (2001). Peer influence groups: identifying dense clusters in large networks. *Social Networks*, 23(4):261–283.

Nivanen, L., Le Mehaute, A., and Wang, Q. (2003). Generalized algebra within a nonextensive statistics. *Reports on Mathematical Physics*, 52(3):437–444.

Park, J. and Newman, M. E. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6):066117.

Pattison, P. and Robins, G. (2002). Neighborhood-based models for social networks. *Sociological Methodology*, 32(1):301–337.

Rényi, A. et al. (1961). On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 547–561.

Riek, R. and Sobol, A. (2016). Comments on the extensivity of the boltzmann entropy. *J Phys Chem Biophys*, 6(207):2161–0398.

Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social networks*, 29(2):192–215.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Schweinberger, M. and Handcock, M. S. (2015). Local dependence in random graph models: characterization, properties and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):647–676.

Sears, T. D. and Sunehag, P. (2007). Induced semantics for undirected graphs: another look at the hammersley-clifford theorem. In *BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING: 27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 954, pages 125–132. AIP Publishing.

Shannon, C. E. (1949). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55.

Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models.

Snijders, T. A. and Baerveldt, C. (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of Mathematical Sociology*, 27(2-3):123–151.

Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153.

Snijders, T. A., Van de Bunt, G. G., and Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social networks*, 32(1):44–60.

Sood, V. and Redner, S. (2005). Voter model on heterogeneous graphs. *Physical review letters*, 94(17):178701.

Touchette, H. (2009). The large deviation approach to statistical mechanics. *Physics Reports*, 478(1):1–69.

Tsallis, C. (1988). Possible generalization of boltzmann-gibbs statistics. *Journal of statistical physics*, 52(1-2):479–487.

Tsallis, C. (2009). Nonadditive entropy: the concept and its use. *The European Physical Journal A-Hadrons and Nuclei*, 40(3):257–266.

Tsallis, C. et al. (2009). *Introduction to nonextensive statistical mechanics*. Springer.

Umarov, S. and Tsallis, C. (2016). The limit distribution in the q-clt for is unique and can not have a compact support. *Journal of Physics A: Mathematical and Theoretical*, 49(41):415204.

Umarov, S., Tsallis, C., and Steinberg, S. (2008). On a q-central limit theorem consistent with nonextensive statistical mechanics. *Milan journal of mathematics*, 76(1):307–328.

Valente, T. W., Fujimoto, K., Chou, C.-P., and Spruijt-Metz, D. (2009). Adolescent affiliations and adiposity: a social network analysis of friendships and obesity. *Journal of Adolescent Health*, 45(2):202–204.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to Markov graphs. *Psychometrika*, 61(3):401–425.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393(6684):440–442.

3 Supplementary Materials (A)

3.1 Power law random graph models

Result 4. *If Assumption 1 is satisfied, then the probability distribution over the networks takes the form*

$$P(\boldsymbol{e}) = \frac{1}{c(\boldsymbol{\theta}, q)} \exp_q\Big(\sum_{k=1}^{p} \theta_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\Big) \tag{11}$$

with $\theta_k = -2h_q(1-c_k)$, $\varphi(\boldsymbol{\theta}) = \sum_{k=1}^{p} N_k \log_q(h_q^{-1}(-\boldsymbol{\theta}_k))$ *and* $h_q(x) = \frac{1}{2}(\log_q(1+x) - \log_q(1-x))$.

Proof of Result 4:

First, let us construct the following probability distribution over the networks satisfying Assumption 1.

$$P(\boldsymbol{e}) = \frac{1}{c(\boldsymbol{\theta}, q)}\Big(\bigotimes_{k=1}^{p} \bigotimes_{\boldsymbol{e}_{S'} \in S_k} Q_{S'}^{(k)}(\boldsymbol{e}_{S'})\Big) \tag{12}$$

where $Q_{S'}^{(k)}(\boldsymbol{e}_{S'}) = \exp_q(\log_q(Q_{S'}^{(k)}(\boldsymbol{e}_{S'})))$ and $c(\boldsymbol{\theta}, q)$ is a normalising constant. To simplify notation, we omit in $\otimes$ the subindex q. Recalling that $\exp_q(a) \otimes_q \exp_q(b) = \exp_q(a+b)$, we have that the right side of the previous equation equals

$$\exp_q\Big(\sum_{k=1}^{p} \sum_{\boldsymbol{e}_{S'} \in S_k} \log_q(Q_{S'}^{(k)}(\boldsymbol{e}_{S'}))\Big) = \exp_q\Big(\sum_{k=1}^{p} \sum_{\boldsymbol{e}_{S'} \in S_k} \big(\frac{c_k^{1-q} - (1-c_k)^{1-q}}{1-q}\big)\mathbb{1}_{\{\boldsymbol{e}_S \sim H_k\}} + \frac{(1-c_k)^{1-q}}{1-q}\Big) \tag{13}$$

θ_k is also equal to $\log_q(c \ominus_q (1-c))$ where $\ominus_q$ is the q-substraction operation defined in Borges (2004).

Equality in (13) follows from

$$\log_q(Q_{S'}^{(k)}(\boldsymbol{e}_{S'})) = \Big(\frac{c_k^{1-q} - (1-c_k)^{1-q}}{1-q}\Big)\mathbb{1}_{\{\boldsymbol{e}_S \sim H_k\}} + \frac{(1-c_k)^{1-q}}{1-q}$$

If we define $h(x) = \log_q(x) - \log_q(1-x)$ and $\theta_k = h(c_k)$, we have that $\frac{\partial h(x)}{\partial x} = (\frac{1}{x})^p + (\frac{1}{1-x})^p > 0$ for all p and $x > 0$. In particular, for all $x \in (0,1)$, h is invertible in $(0,1)$. Thus, the right side of Equation (13) can be rewritten as follows

$$\exp_q\Big(\sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'} \in S_k}(\theta_k \mathbb{1}_{\{\boldsymbol{e}_S \sim H_k\}}) + \sum_{k=1}^{p} N_k \log_q(h^{-1}(\theta_k))\Big)$$

By defining $\Gamma_k(\boldsymbol{e}) = \sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'} \in S_k} \mathbb{1}_{\{\boldsymbol{e}_S \sim H_k\}}$ and $\varphi(\boldsymbol{\theta}) = \sum_{k=1}^{p} N_k \log_q(h^{-1}(\theta_k))$, we have that the last equation equals Equation (11).

■

The derivatives of the normalising constant $c(\boldsymbol{\theta}, q)$ and $\varphi(\boldsymbol{\theta})$ satisfy the following properties

$$\frac{\partial \log(c(\boldsymbol{\theta}, q))}{\partial \theta_k} = \frac{v_q(P_{\boldsymbol{\theta}})}{v_1(P_{\boldsymbol{\theta}})}\sum_{\tilde{\boldsymbol{e}}}\Big(\Gamma_k(\tilde{\boldsymbol{e}}) + \frac{\partial \varphi(\boldsymbol{\theta})}{\partial \theta_k}\Big)P_{\boldsymbol{\theta},q}(\tilde{\boldsymbol{e}}) = \frac{v_q(P_{\boldsymbol{\theta}})}{v_1(P_{\boldsymbol{\theta}})}\Big(\mathrm{E}_{\boldsymbol{\theta},q}(\Gamma_k(\tilde{\boldsymbol{e}})) + \frac{\partial \varphi(\boldsymbol{\theta})}{\partial \theta_k}\Big)$$

With $v_q(P_{\boldsymbol{\theta}}) = \sum P_{\boldsymbol{\theta}}^q(\tilde{\boldsymbol{e}})$ and $P_{\boldsymbol{\theta},q}(\tilde{\boldsymbol{e}}) = \frac{P_{\boldsymbol{\theta}}^q(\tilde{\boldsymbol{e}})}{v_q(P_{\boldsymbol{\theta}})}$.

$$\frac{\partial \varphi(\boldsymbol{\theta})}{\partial \theta_k} = N_k \frac{c_k^q}{c_k^q + (1-c_k)^q}$$

Consistent with results obtained for the ERGM, we have for $q = 1$,

$$\frac{\partial \varphi(\boldsymbol{\theta})}{\partial \theta_k} - \frac{\partial \log(c(\boldsymbol{\theta}, 1))}{\partial \theta_k} = -\mathrm{E}_{\boldsymbol{\theta}}(\Gamma_k(\tilde{\boldsymbol{e}}))$$

A second representation of the probability distribution shown in Equation (12) is

$$P(\boldsymbol{e}) = (\frac{1}{\tilde{\boldsymbol{c}}}) \otimes_q \Big(\bigotimes_{k=1}^{p}\bigotimes_{\boldsymbol{e}_{S'} \in S_k} \tilde{Q}_{S'}^{(k)}(\boldsymbol{e}_{S'})\Big) \tag{14}$$

with $\tilde{Q}^{(k)}_{S'}(\boldsymbol{e}_{S'}) = \exp_q\Big(\big(\frac{1}{c_{ij}}\big)^{1-q}\log_q\big(Q^{(k)}_{S'}(\boldsymbol{e}_{S'})\big)\Big)$ and for some constant $\tilde{\boldsymbol{c}}$. This representation is important for linking PRGM with q-Markov random fields.

The new representation follows by noticing that Equation (12) equals

$$P(\boldsymbol{e}) = \frac{1}{c(\boldsymbol{\theta}, q)}\exp_q\Big(\sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'}\in S_k}\log_q\big(Q^{(k)}_{S'}(\boldsymbol{e}_{S'})\big)\Big) \tag{15}$$

Next, notice that the q-exponential function satisfies the following property:

$$a\exp_q(x) = \exp_q\big(a^{1-q}x + \frac{a^{1-q}-1}{1-q}\big) \tag{16}$$

From Equation (15) and (16) follows

$$P(\boldsymbol{e}) = \exp_q\Big(\sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'}\in S_k}\big(\frac{1}{c(\boldsymbol{\theta}, q)}\big)^{1-q}\log_q\big(Q^{(k)}_{S'}(\boldsymbol{e}_{S'})\big) + \log_q\big(\frac{1}{c(\boldsymbol{\theta}, q)}\big)\Big) \tag{17}$$

Finally using the property $\exp_q(x+y) = \exp_q(x)\otimes_q \exp_q(y)$, defining $\tilde{Q}^{(k)}_{S'}(\boldsymbol{e}_{S'}) = \exp_q\Big(\big(\frac{1}{Z}\big)^{1-q}\log_q\big(Q^{(k)}_{S'}(\boldsymbol{e}_{S'})\big)\Big)$ and $\tilde{\boldsymbol{c}} = \frac{1}{\exp_q(\log_q(\frac{1}{Z}))}$, we have that Equation (17) equals Equation (14).

3.2 q-Markov random field

Result 5. *For any PRGM, its dependency graph G_q satisfies the pairwise and local Markov property. When $q = 1$, the global property also holds.*

Proof of Result 5: (q**-pairwise Markov property**)

Let e_{ij} and $e_{i'j'}$ be two variables not connected in G_q. Next, notice that there is no function $Q_{S'}$ having in their domain both variables e_{ij} and $e_{i'j'}$. If this were not the case, then both variables would be connected in G_q, which is a contradiction. By using the commutative and associativity of the q-product, we can rewrite Equation (14) as follows:

$$P(\boldsymbol{e}) = \overbrace{(\frac{1}{\tilde{\boldsymbol{c}}}) \bigotimes (\bigotimes_{k=1}^{p} \bigotimes_{\substack{e_{ij} \in \boldsymbol{e}_S \\ \boldsymbol{e}_S \in S_k}} \tilde{Q}_S(\boldsymbol{e}_S))}^{f_1(\boldsymbol{e}_{-i'j'})} \bigotimes \overbrace{(\bigotimes_{k=1}^{p} \bigotimes_{\substack{e_{ij} \notin \boldsymbol{e}_S \\ \boldsymbol{e}_S \in S_k}} \tilde{Q}_S(\boldsymbol{e}_S))}^{f_2(\boldsymbol{e}_{-ij})} \tag{18}$$

By conditioning on the set of variables $\boldsymbol{e}_{-\{ij,i'j'\}}$, we have that

$$P(e_{ij}, e_{i'j'} | \boldsymbol{e}_{-\{ij,i'j'\}}) = \frac{1}{c}(f_1(e_{ij}|\boldsymbol{e}_{-\{ij,i'j'\}}) \otimes_q f_2(e_{i'j'}|\boldsymbol{e}_{-\{ij,i'j'\}}) \tag{19}$$

where c is the marginal density for $\boldsymbol{e}_{-\{ij,i'j'\}}$.

$$c = \sum_{(e_{ij}, e_{i'j'}) \in \{0,1\}^2} P(e_{ij}, e_{i'j'}, \boldsymbol{e}_{-\{ij,i'j'\}})$$

Finally, applying Formula (16), we have that

$$P(e_{ij}, e_{i'j'} | \boldsymbol{e}_{-\{ij,i'j'\}}) = \tilde{f}_1(e_{-i'j'}) \otimes_q \tilde{f}_2(e_{-ij}) \tag{20}$$

The proof for q-local Markov property follows the same steps.

■

Definition 8. *We say that the q-dependency graph G satisfies the* ***weak q-Global Markov property*** *if and only if for any partition of the variables $\boldsymbol{e}_S$ of the form $\boldsymbol{e}_S = \boldsymbol{e}_A \cup \boldsymbol{e}_B \cup \boldsymbol{e}_C$ with $\boldsymbol{e}_A$ and $\boldsymbol{e}_B$ separated by $\boldsymbol{e}_C$, we have that*

$$\boldsymbol{e}_A \perp\!\!\!\perp_q \boldsymbol{e}_B | \boldsymbol{e}_C$$

Result 6. *Any PRGM satisfies the weak q-global Markov property.*

Proof of Result 6:

First, let us notice that for any $Q_{S'}$, we have that either $\boldsymbol{e}_{S'} \cap \boldsymbol{e}_A = \emptyset$ or $\boldsymbol{e}_{S'} \cap \boldsymbol{e}_B = \emptyset$. If this were not the case, then there would exist a link between some variables in $\boldsymbol{e}_A$ and some variables in $\boldsymbol{e}_B$. Let $[\boldsymbol{e}_A]$ be defined as as follows: for any $Q_{S'}$ $\boldsymbol{e}_{S'} \in [\boldsymbol{e}_A]$ if and only if $\boldsymbol{e}_{S'} \cap \boldsymbol{e}_A \neq \emptyset$. Similarly, we define $[\boldsymbol{e}_B]$ and $[\boldsymbol{e}_C]$ is defined as the set of $\boldsymbol{e}_{S'}$ not contained in $[\boldsymbol{e}_A] \cup [\boldsymbol{e}_B]$ Next, we rewrite Equation (11) as follows

$$P(\boldsymbol{e}) = \frac{1}{Z} \bigotimes_{k=1}^{p} \bigotimes_{\substack{\boldsymbol{e}_S \in [\boldsymbol{e}_A] \\ \boldsymbol{e}_S \in S_k}} Q_S(\boldsymbol{e}_S) \bigotimes_{k=1}^{p} \bigotimes_{\substack{\boldsymbol{e}_S \in [\boldsymbol{e}_B] \\ \boldsymbol{e}_S \in S_k}} Q_S(\boldsymbol{e}_S) \bigotimes_{k=1}^{p} \bigotimes_{\substack{\boldsymbol{e}_S \in [\boldsymbol{e}_C] \\ \boldsymbol{e}_S \in S_k}} Q_S(\boldsymbol{e}_S) \tag{21}$$

Conditioning on $\boldsymbol{e}_C$ and following the same arguments as for the q-parwaise case, we have that $P(\boldsymbol{e}_A, \boldsymbol{e}_B | \boldsymbol{e}_C) = f_1(\boldsymbol{e}_A|\boldsymbol{e}_C) \otimes_q f_2(\boldsymbol{e}_B|\boldsymbol{e}_C)$

If we take a $\boldsymbol{e}_{B'} \subset \boldsymbol{e}_B$ and $q = 1$ and using the property (a) $\sum_x \prod b f(x,y) = b \prod \sum_x f(x,y)$, we have that

$$P(\boldsymbol{e}_A, \boldsymbol{e}_{B'} | \boldsymbol{e}_C) = f_1(\boldsymbol{e}_A) f_2(\boldsymbol{e}_{B'})$$

However, property (a) does not hold for $q \neq 1$ and we cannot use this step for the general case.

■

3.3 Power law random graph model and Tsallis entropy

Result 7. *Let*

$$H_q(\boldsymbol{e}) = \sum_{\boldsymbol{e}\in\mathcal{G}^n} P(\boldsymbol{e}) \log_q\Big(\frac{1}{P(\boldsymbol{e})}\Big) \tag{22}$$

Subject to

$$\sum_{\boldsymbol{e}\in\mathcal{G}^n} P(\boldsymbol{e}) = 1 \tag{23}$$

and

$$\sum_{\boldsymbol{e}\in\mathcal{G}^n} \Gamma_k(\boldsymbol{e}) P_q(\boldsymbol{e}) = \bar{\Gamma}_k \tag{24}$$

for some constants $\bar{\Gamma}_k$ *and*

$$P_q(\boldsymbol{e}) = \frac{P^q(\boldsymbol{e})}{\sum_{\boldsymbol{e}\in\mathcal{G}^n} P^q(\boldsymbol{e})}$$

If the solution to the Maximization problem exists, then it has the form

$$P(\boldsymbol{e}) = \frac{1}{c(\boldsymbol{\theta},q)}\Big(1+(1-q)\big(\varphi(\boldsymbol{\theta})-\sum_{k=1}^{p}\theta_k\Gamma_k(\boldsymbol{e})\big)\Big)^{\frac{1}{1-q}} = \frac{1}{c(\boldsymbol{\theta},q)}\exp_q\Big(\sum_{k=1}^{p}\theta_k\Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\Big) \tag{25}$$

for some $\boldsymbol{\theta} = (\theta_1, ..., \theta_p) \in \mathbb{R}^p$*,* $\varphi(\boldsymbol{\theta}) =< \boldsymbol{\theta}, \bar{\boldsymbol{\Gamma}} >$ *and* $c(\boldsymbol{\theta}, q)$ *a normalising constant.*

Before proving Result (7), let us rewrite the optimisation problem as follows:

Maximize

$$\frac{1}{1-q}\Big(\sum_{\boldsymbol{e}\in\mathcal{G}^n} P^q(\boldsymbol{e}) - 1\Big)$$

Subject to

$$\sum_{\boldsymbol{e}\in\mathcal{G}^n} P(\boldsymbol{e}) = 1 \tag{26}$$

$$\sum_{\boldsymbol{e}\in\mathcal{G}^n} \Gamma_k(\boldsymbol{e})P^q(\boldsymbol{e}) = \bar{\Gamma}_k\Big(\sum_{\boldsymbol{e}\in\mathcal{G}^n} P^q(\boldsymbol{e})\Big) \tag{27}$$

Proof of Result 7:

The solution to the optimisation problem is done by introducing the Lagrange multipliers $\{\theta_i\}_{i=1}^p$ and the Lagrange function:

$$\mathcal{L} = \frac{1}{1-q}\Big(\sum_{\boldsymbol{e}} P^q(\boldsymbol{e}) - 1\Big) - \theta_0(\sum_{\boldsymbol{e}} P(\boldsymbol{e})) - \sum_{k=1}^{p}\theta_k \sum_{\boldsymbol{e}} P^q(\boldsymbol{e})(\Gamma_k(\boldsymbol{e}) - \bar{\Gamma}_k)$$

Taking the partial derivatives of $\mathcal{L}$, we find that the maximum entropy is achieved for the distribution satisfying

$$\frac{1-q}{q}P^{q-1}(\boldsymbol{e}) - \theta_0 - q\sum_{k=1}^{p}\theta_k P^{q-1}(\boldsymbol{e})(\Gamma_k(\boldsymbol{e}) - \bar{\Gamma}_k) = 0$$

$$P(\boldsymbol{e}) = \big(\frac{(1-q)\theta_0}{q}\big)^{\frac{1}{1-q}}\Big(1 - (1-q)\sum_{k=1}^{p}\theta_k\big(\Gamma_k(\boldsymbol{e}) - \bar{\Gamma}_k\big)\Big)^{\frac{1}{1-q}}$$

$$P(\boldsymbol{e}) = \frac{1}{\boldsymbol{c}}\exp_q\big(-\sum_{k=1}^{P}\theta_k\Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big) \tag{28}$$

with $\boldsymbol{c} = \big(\frac{(1-q)\theta_0}{q}\big)^{\frac{-1}{1-q}}$ and $\varphi(\boldsymbol{\theta}) = \sum_{k=1}^{p}\theta_k\bar{\Gamma}_k$

From the constraints we have that

$$\boldsymbol{c} = \sum_{\boldsymbol{e}} \exp_q\big(-\sum_{k=1}^{p} \theta_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big)$$

■

For $q = 1$, Formula (28) equals

$$P(\boldsymbol{e}) = \exp\big(-\sum_{k=1}^{p} \theta_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta}) - \log(\frac{1}{Z})\big)$$

and $\frac{\partial \varphi(\boldsymbol{\theta}) - \log(\frac{1}{Z})}{\partial \theta_k} = \mathrm{E}_{\boldsymbol{\theta}}(\Gamma_k)$.

3.4 An Implication of large q

For the q-two star and q-triangle models, it was observed that when q is larger than 1, the probability distributions tends to depart from distributions placing most of their mass around the null- and complete networks. This follows from the fact that as q grows and the parameter $\boldsymbol{\theta}$ is held constant, the forces of the social mechanisms driving the system to these two extreme cases are shadowed by the idiosyncratic correlation. The conditional odds of a relation between agent i and j, denoted by $odds_{ij}$, tends to one as q tends to infinity.

$$odds_{ij} = \frac{P(e_{ij} = 1|\boldsymbol{e}_{-ij})}{P(e_{ij} = 0|\boldsymbol{e}_{-ij})} = \Big(\frac{1 + (1-q)\big(\varphi(\boldsymbol{\theta}) - \sum_{k=1}^{p} \theta_k \Gamma_k(e_{ij} = 1, \boldsymbol{e_{-ij}})\big)}{1 + (1-q)\big(\varphi(\boldsymbol{\theta}) - \sum_{k=1}^{p} \theta_k \Gamma_k(e_{ij} = 0, \boldsymbol{e_{-ij}})\big)}\Big)^{\frac{1}{1-q}}$$

Since for any positive number a, we have that $\lim_{x\to\infty} a^{\frac{-1}{x}} = 1$, it follows that

$$\lim_{q\to\infty} odds_{ij} = 1$$

3.5 q-exponential random graph models

Motivated by the derivation of the exponential random graph models by (Wasserman and Pattison, 1996) and the q-exponential family of probability distributions (Amari and Ohara, 2011), we present a class of network models closely related to the PRGM. We call this class of network models q-exponential random graph models (q-ERGM).

$$P_{\boldsymbol{\theta}}(\boldsymbol{e}_S) = \exp_q\Big(-\sum_{k=1}^{p}\theta_k\Gamma_k(\boldsymbol{e}_S) + \varphi(\boldsymbol{\theta})\Big) \tag{29}$$

Γ_k represents a network statistic and $\varphi(\boldsymbol{\theta})$ is a normalising constant. For the q-exponential random graph, estimation of the parameter $\boldsymbol{\theta}$ is simplified by working with the q-maximum likelihood estimator, $l_q(\boldsymbol{\theta})$.

Contrary to PRGM, the q-ERGM do not find its foundation at the agent-agent interaction level.

$$l_q(\boldsymbol{\theta}) = -\sum_{k=1}^{p}\theta_k\Gamma_k + \varphi(\boldsymbol{\theta}) \tag{30}$$

$$\frac{\partial l_q(\boldsymbol{\theta})}{\partial\theta_k} = -\Gamma_k + \frac{\partial\varphi(\boldsymbol{\theta})}{\partial\theta_k} \tag{31}$$

Result 8.

$$\frac{\partial\varphi(\boldsymbol{\theta})}{\partial\theta_k} = \mathrm{E}_{\boldsymbol{\theta},q}(\Gamma_k(\boldsymbol{e})) = \sum_{\boldsymbol{e}\in\mathcal{G}^n}\Gamma_k(\boldsymbol{e})P_q(\boldsymbol{e}) \tag{32}$$

This approach does not makes assumption at the agent-agent interaction level, but postulates that the distribution over the network is defined by known network statistics.

$$\frac{\partial^2\varphi(\boldsymbol{\theta})}{\partial^2\theta_k} = q\mathrm{E}_{\boldsymbol{\theta},q}\Big((\Gamma_k(\boldsymbol{e}) - \mathrm{E}_{\boldsymbol{\theta},q}(\Gamma_k))^2\Big) = q\mathrm{Var}_{\boldsymbol{\theta},q}\Big(\Gamma_k(\boldsymbol{e})\Big) \tag{33}$$

$$\frac{\partial l_q(\boldsymbol{\theta})}{\partial\theta_k} = 0 \quad \text{if and only if} \quad \Gamma_k(\boldsymbol{e}) = \mathrm{E}_{\boldsymbol{\theta},q}(\boldsymbol{e}) \tag{34}$$

Proof of Result 8:

Part (1): Let us consider the constraint

$$\sum_{\boldsymbol{e}\in\mathcal{G}^n} \exp_q\big(- \sum_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big) = 1$$

Deviating with respect to θ_k on both sides

$$\sum_{\boldsymbol{e}\in\mathcal{G}^n} \Big(- \Gamma_k(\boldsymbol{e}) + \frac{\partial\varphi(\boldsymbol{\theta})}{\partial\theta_k}\Big) \exp_q^q\big(- \sum_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big) = 0 \tag{35}$$

which implies

$$\frac{\partial\varphi(\boldsymbol{\theta})}{\partial\theta_k} = \frac{1}{\sum_{\boldsymbol{e}\in\mathcal{G}^n} \exp_q^q\big(- \sum_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big)} \sum_{\boldsymbol{e}\in\mathcal{G}^n} \Gamma_k(\boldsymbol{e}) \exp_q^q\big(- \sum_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big)$$

Part (2): Taking the second derivative of Equation (35) gives

$$\sum_{\boldsymbol{e}\in\mathcal{G}^n} \frac{\partial^2\varphi(\boldsymbol{\theta})}{\partial\theta_k^2} \exp_q^q\big(- \sum_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big) + \sum_{\boldsymbol{e}\in\mathcal{G}^n} q\Big(-\Gamma_k(\boldsymbol{e}) + \frac{\varphi(\boldsymbol{\theta})}{\partial\theta_k}\Big)^2 \exp_q^q\big(- \sum_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big) = 0$$

which implies

$$\frac{\partial^2 \varphi(\boldsymbol{\theta})}{\partial \theta_k^2} = \frac{1}{\sum_{\boldsymbol{e}\in\mathcal{G}^n} \exp_q^q\big(- \sum_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big)} \sum_{\boldsymbol{e}\in\mathcal{G}^n} q\Big(-\Gamma_k(\boldsymbol{e}) + \frac{\varphi(\boldsymbol{\theta})}{\partial\theta_k}\Big)^2 \exp_q^q\big(- \sum_k \Gamma_k(\boldsymbol{e}) + \varphi(\boldsymbol{\theta})\big)$$

■

4 Revealing the effects of network size on social mechanisms

Abstract

Which are the mechanisms that underlie the formation of networks? This is a key question in network science, pervading the most variegated disciplines, and extensively approached theoretically and empirically. Importantly, most results build on assuming a large number of network constituents, while little is known for systems where this is implausible. For statistical network analysis, extant methodologies have two shortcomings: (1) parameters are not constant on network size and (2) many exemplary social networks consist of just a few dozens of agents. We address the first problem by determining the functional form of the parameters on the network size for parsimonious exponential random graph models. We use these results to construct a new class of models, termed *finite exponential random graph model* (fERGM) which do not make presuppositions on the network size, but resort on a sample of observed networks. This exchange of premises is necessary for analyzing small network,s and it allows to study the influence of size on the network formation. Further, we demonstrate how to use fERGM to test for the effect that the network size has on simple mechanisms for the formation of networks.

4.1 Introduction

In most social processes, the decisions of an individual are not just the result of internal reflections but are influenced by the choices of others (Goodreau et al., 2009) and by the social environment in which the agents coexist (Guo and Zhao, 2000). For example, the tendency of agents to reciprocate friendship ties and favors (Vaquera and Kao, 2008) or to become friends with a mutual friend (Louch, 2000) are examples of the former type of influence, and they have been extensively studied in the social networks literature (Wasserman and Faust, 1994). Problems such as the effect of classroom size on students' performance (Ehrenberg et al., 2001) or the consequences of neighborhood poverty on delinquent behavior (Oberwittler, 2007) are just some examples of the latter and are the focus of an extensive body of research in the social sciences (Bryk and Raudenbush, 1992). Extant research on social networks has acknowledged the importance that the environment has on the formation of networks (Lubbers, 2003; McFarland et al., 2014; Lazega and Snijders, 2015). However, current methodologies for cross-sectional

Author Statement: This work has been done together with René Algesheimer and Claudio Tessone. A version of this paper has been submitted to the journal Network Science.

network data do not allow to test for existence - nor to quantify the actual effect - of the environment on the formation of links in small networks. Fundamental, yet simple, research questions such as: Are the social mechanisms underlying the formation of network influenced by the number of agents constituting the network.

A main reason is that the explanation to phenomena commonly found in empirical networks rely on some kind of asymptotic limit (at variegated levels of network description), which only becomes valid as the number of agents is set to diverge. Examples of these include: the concept of almost all graphs (Erdős and Rényi, 1960), the emergence of power-law degree distributions (Barabási and Albert, 1999) and the small-world phenomenon (Watts and Strogatz, 1998), just to cite the most paradigmatic ones. Therefore, well-documented network properties in large networks cannot simply be assumed to exist in e.g. social networks where the number of agents can be of the order of tens or hundreds of agents.

Hypothesis testing of network formation processes offers an example for a situation, in which assuming large network size severely limits the applicability of extant methods. Historically, this has been done with a variety of models, ranging from those that assume independence between links to others that take into account their natural interdependence (Frank and Strauss, 1986; Robins et al., 2007). Among the latter, exponential random graph models (ERGM) have gained particular prominence (Lusher et al., 2012). By testing hypotheses on the parameters of ERGMs, researchers have claimed to identify the existence of reciprocity, transitivity and homophily in a variety of empirical networks (Wu et al., 2015; Kim et al., 2015; Jasny et al., 2015). Commonly, these analyses have been performed on a single network, and the tests have been constructed by imposing assumptions at the agent-agent interaction level (Chandrasekhar and Jackson, 2014; Schweinberger and Handcock, 2015) or by introducing implicit assumptions (Hunter and Handcock, 2006; Cranmer and Desmarais, 2011). These assumptions are then used to justify some asymptotic properties when the number of agents tends to infinity, e.g. the limiting distribution of network statistics becomes Gaussian. However, even simple agent to agent dependencies may cause non-Gaussian distributions of the test statistics, which may in-validate conclusions of statistical network analyses, e.g. bimodal distributions (Handcock et al., 2003). Another shortcoming of this approach becomes evident if one aims at studying the effect of the number of agents on the network formation process.

Some recent papers have tested for the effect of environmental variables on the formation of networks by analyzing multiple networks (Lubbers, 2003; Valente et al., 2009; Huitsing and Veenstra, 2012; Schaefer and Simpkins, 2014; Song, 2015; Kruse et al., 2016). However, in this article, we show that shortcomings subsist. First, the most predominant methodology for analysing multiple networks with cross-sectional data assumes that each observed network consists of multiple observations (Lubbers, 2003), which does not necessarily hold for small networks and when agents' decisions are strongly correlated. Further, we show that the prevailing methodology use estimators which are not consistent as the number of observed networks tends to infinity. Second, it requires the knowledge of the functional form of the ERGM parameters on the number of agents, which has not been well understood. The three main contributions of this Paper can be summarized as follows. First, we introduce a new class of models that: (i) do not assume that the number of agents tend to infinity, (ii) can be used to simultaneously analyze multiple networks of different sizes generated by the same linking formation process. We call this approach finite exponential random graph model, or fERGM for short. fERGM complements previous statistical network models because it replaces assumptions at the agent-agent interaction level by milder assumptions on the data generating process. Second, we provide the functional form of parameters needed for simple mechanisms of network formation. Finally, we show how to use fERGM to test the effects that the number of agents has on the formation of networks for two central processes, i.e. reciprocity and transitivity. We achieve this by analyzing 214 directed networks from four different datasets.

4.2 Network Model

The goals of this section are three-fold: (1) show that ERGM parameters are under weak assumptions a non-constant function of network size and to give some approximations for these functions under certain conditions, (2) to describe issues of current statistical network research based on ERGM and multilevel network analysis combining two step procedures and ERGM, (3) to introduce the finite exponential random graph model (*f*ERGM).

4.2.1 Exponential random graph model (ERGM)

Although observed networks can be modelled by a probability distribution over a network *ensemble*, it is more natural to think about them as the outcome of the formation and severing of links between agents. Therefore, we first model the updating process of links as conditional probabilities; and only then we present the relation between the conditional probabilities obtained and ERGM, e.g. probability distribution over networks.

We assume that the decision of an agent to form or sever a link with another depends on properties this link (or its absence) will induce at the local network level. Formally, we consider there is a finite set of agents $V = \{1, ..., n\}$, each agent $i \in V$ is assigned a set of indicator relational variables $\{e_{ij} \mid j \in V \setminus \{i\}\}$ with e_{ij} equal 1 if and only if there is a link (relation) from agent i to j. We denote $\boldsymbol{e}$, $\boldsymbol{e}_{-ij}$ and $\boldsymbol{e}_S$ the set of all relational variables, the set of all relational variables except for e_{ij} and a subset of relational variables, respectively. $G = G[\boldsymbol{e}_S]$ denotes the induced subnetwork with links defined by $\boldsymbol{e}_S$ and agents being endpoints of the relational variable in $\boldsymbol{e}_S$. Two subnetworks G, $\tilde{G}$ are said to be isomorphic if there is a bijection f between the agents that preserves existing and non-existing relations, i.e. there is (not) a link from i to j in G if and only if there is (not) a link from $f(i)$ to $f(j)$ in $\tilde{G}$. Notice that by definition the induced subnetwork given by $\boldsymbol{e}_S = \{e_{ij}\}$ and $e_{ij} = 1$ is not isomorphic to the induced subnetwork given by $\boldsymbol{e}_{S'} = \{e_{ij}, e_{ji}\}$ and $e_{ij} = 1, e_{ji} = 0$. In the first subnetwork we do not allow agent j to form a link to i while in the second one agent j is allowed to choose to form a link and the choice is not to.

A *structure* is a set of subnetworks isomorphic to each other, and when a subnetwork is in a structure we say that the subnetwork has that structure. Given a set of relational variables $\boldsymbol{e}_S$ agents often prefer to update their relations such that they are in a subnetwork with a certain structure. For instance, in the set of relational variables $\{e_{ij}\}$, it might happen that agent i prefers to be related to j, in which case he prefers to be in a subnetwork with a structure defined by two agents and e_{ij} being equal to one. Next, we assume that the probability to update a relation is driven by a class of tendencies to create subnetworks with certain structures. We denote the class of these structures by the index $\mathcal{I}$ ($|\mathcal{I}| < \infty$), and S_k is the set of subsets of relational variables such that if $\boldsymbol{e}_S \in S_k$, then there exists a realisation $\bar{e}_S$ such that the induced graph has structure H_k with $k \in \mathcal{I}$.

The conditional probability distribution of a relational variable e_{ij} on the remaining relational variables is

$$P(e_{ij} \mid \boldsymbol{e}_{-ij}) = c_{ij} \prod_{k \in \mathcal{I}} \prod_{\substack{e_{ij} \subset \boldsymbol{e}_S \\ \boldsymbol{e}_S \in S_k}} Q_S(\boldsymbol{e}_S), \tag{36}$$

with c_{ij} a constant. In simple terms, this set of equations states that the conditional distribution of relational variables can be partitioned into different blocks, each one representing the tendency to create a subnetwork with a concrete structure.

Further, we state that $Q_S(\boldsymbol{e}_S) = c_k$ if $G[\boldsymbol{e}_S] \sim H_k$, and $Q_S(\boldsymbol{e}_S) = 1 - c_k$, otherwise for some $c_k \in (0,1) \setminus \{\frac{1}{2}\}$. When $c_k > 1/2$, this tendency states that agents are more likely to update e_{ij} in such a way that it is in a subnetwork with structure H_k, and for $c_k < 1/2$ agents are more likely to update e_{ij} such that it is not in a subnetwork with structure H_k. If updating e_{ij} cannot contribute in adding a subnetwork with structure H_k, this tendency does make more or less likely to add a link. For example, if H_l denotes the structure defined by two agents $\{i, j\}$ and the link from i to j and H_r the structure defined by two agents and the links from i to j and from j to i; then H_l might model the tendency of agents to relate with other and H_r might model the tendency to reciprocate a relation.

Given a network $G = G[\boldsymbol{e}]$ induced by the set of agents $\boldsymbol{e}$, the set of Eqs. (36) defines the exponential random graph model (see SI), i.e.

$$P_{\boldsymbol{\theta}}(G) = \frac{1}{c(\boldsymbol{\theta})} \exp\left(\sum_{k\in\mathcal{I}} \theta_k \Gamma_k(G)\right). \tag{37}$$

Here, $\theta_k = \log(\frac{c_k}{1-c_k})$, $\Gamma_k(G)$ is a *network statistics*, equal to the number of subnetworks in G isomorphic to a structure H_k, and $c(\boldsymbol{\theta})$

$$c(\boldsymbol{\theta}) = \sum_{G\in\mathcal{G}^n} \exp\left(\sum_{k\in\mathcal{I}} \theta_k \Gamma_k(G)\right).$$

where the dimension of $\boldsymbol{\theta}$ equals the cardinality of $\mathcal{I}$.

In formal terms, a *social mechanism* is the tendency of agents to belong to a subnetwork of certain structures and denoted by the index $\mathcal{I}$. Then, a *linking process* is that of randomly selecting a pair of agents i, j and the decision of agent i to update links according to a social mechanism. A set of social mechanisms has to be provided when using ERGM to model the formation of a network. Some examples of social mechanisms are (i) agents tendency to relate to others; the concomitant network statistics for it is the total number of links (*l*), with associated parameter θ_l. (ii) agents' tendency to reciprocate a relation; in this case the network statistics is the number of reciprocal pairs of links (*r*) with associated parameter θ_r; (iii) agent's tendency to relate to a friend's friend; this (more involved) property can be seen as the combination of the tendency of agents to be in a transitive triangle but not in a mixed two-path. Therefore, two additional network statistics are needed: the number of transitive triangles *t* and the number of mixed two paths *m2p* (with associated parameters θ_t, θ_{m2p}, respectively). The last modelling step in the use of ERGM is relating the probabilities to create links with certain structures to the social mechanisms.

4.2.2 Properties of parameters

Similar to logistic regressions, it is possible to interpret the parameters with the log odds, i.e.

$$\text{logit}\big(P_{\boldsymbol{\theta}}(e_{ij} = 1|e_{-ij} = \bar{e}_{-ij})\big) = \sum_{k\in\mathcal{I}} \theta_k. \tag{38}$$

For the sake of an example, it is possible to interpret θ_l as the log odds ratio of adding a link while holding constant the other network statistics included in the model. However, a sharp difference is that the parameters are not a constant function of network size, as it is usually expected in regression models.

To understand this, we have first to notice that the larger the number of agents, the less likely two agents would *meet*, i.e. be together selected of a link update. Thereby, the less likely the two agents will add a link between them, and intuitively we would expect θ_l to be of the form $\theta_{1,l} + \theta_{2,l} f_l(n)$ for some parameters $\theta_{1,l} \in \mathbb{R}$, $\theta_{2,l} < 0$ and $f_l(n)$ a strictly increasing function of the network size.

Notice that for the simplest ERGM defined only by the network statistics number of links *l*, assuming $f_l(n) \propto \log(n)$ and $\theta_{2,l} = -1$ implies that the expected number of links grows linearly on n. The reason is that the expected number of links in undirected networks is equal to

$$E(e) = \binom{n}{2} p_n = \binom{n}{2} \frac{\exp\left(-\log(n)\right)}{1 + \exp\left(-\log(n)\right)} \propto n,$$

with p_n the probability to have a link between two agents (a similar argument holds for directed networks). Thus, if we expect the number of links to grow approximately linear on the number of agents, then a reasonable assumption is to take $f_l(n) = \log(n)$ and $\theta_{2,l}$ unknown. Stronger assumptions were suggested when the expected average degree converges, $f_l(n) = \log(n)$ and $\theta_{2,l} = -1$ (Krivitsky et al., 2011). Such functional dependency might be reasonable in social networks where links are generated asking agents "name up to k persons you are related with".

In the extreme case when everyone names k different agents, the expected number of links is a linear function of the network size.

A more involved case is a model that includes the network statistics the number of links and number of reciprocal pairs of links, and possibly other network statistics. The log odds ratio of reciprocating a link is now

$$\theta_l + \theta_r \propto \theta_{1,l} + \theta_{1,r} + \theta_{2,l} f_l(n) + \theta_{2,r} f_r(n) \tag{39}$$

when holding all network statistics constant except for l and r. If the probability that an agent adds a link to another one is inversely proportional to the network size and the probability to reciprocate a link is constant, then,

$$f_l(n) = f_r(n), \;\; \theta_{2,l} < 0 \text{ and } \;\; \theta_{2,r} = -\theta_{2,l} > 0.$$

In this case, the probability to reciprocate a link is equal to

$$\frac{\exp\left(\theta_{1,l} + \theta_{1,r}\right)}{1 + \exp(\theta_{1,l} + \theta_{1,r})}$$

when holding all network statistics constant except for *l* and *r*.

If, in addition, the number of edges has a linear growth then by following the same arguments as above, we can conclude that $f_r(n) = \log(n)$. Using similar arguments, we might assume that if transitivity is constant, then $f_t(n) = f_l(n)$ and $\theta_{2,t} = -\theta_{2,l}$, and $f_t(n) = \log(n)$ when there is a linear growth of the number of links. Thus, under some assumptions on the functions f, it is possible to test for constant reciprocity by testing if $\theta_{2,r} = -\theta_{2,l}$ (see SI).

4.2.3 Statistical analysis on multiple networks and finite exponential random graph (fERGM)

Statistical methods for analysing multiple networks were introduced to address the group-composition effects on the formation of networks (Lubbers, 2003), The methodology used estimates parameter on each observed network, as it is built on meta-analyses methods that presuppose that each network consists of several observations (Snijders and Baerveldt, 2003; Lubbers, 2003). This assumption is undermined by common estimation problems in empirical studies (Lubbers, 2003; Goodreau et al., 2009; Lewis, 2013; Boda and Néray, 2015; Kruse et al., 2016), which may be interpreted as the symptom of a small sample size problem caused by the non-independence of the relational variables (Camacho Guardian, 2016). The estimation problem has forced researchers to modify their initial models (Boda and Néray, 2015) or to remove networks from the analyses (Lubbers, 2003; Goodreau et al., 2009; Lewis, 2013; Kruse et al., 2016), and the problem does not fade as the number of collected networks tends to infinity (see SI).

To overcome the deficiencies of existing procedures and by considering each network as a single observation, we introduce the following model. The first modelling step is assuming that there exists an infinite population of agents I, which is randomly partitioned into disjoint sets V_i. p_n denotes the probability that V_i is composed of exactly n agents and the set of all networks with n agents is $\mathcal{G}^n$. Notice that our assumption of an infinite population is quite different from the one assumed in ERGM for consistency and asymptotic normality, as we do not require to observe a single large population but a large sample of observed networks. Second, we assume that inside each element i of the partition the same linking process takes place. Thus, the probability of observing a particular network G is decomposed in our model in two parts: The first is the probability to observe a network with n agents and the second is the probability to observe a network with certain characteristics given that it has a size n.

In formal terms, a *finite exponential random graph model* (fERGM) is a probability space

$(\mathcal{G}^\infty, \mathcal{A}, P_{\boldsymbol{\theta}})$ with $\mathcal{G}^\infty = \cup_{n=1}^{\infty}\mathcal{G}^n$, $\mathcal{A}$ is the class of subsets of $\mathcal{G}^\infty$, and $P_{\boldsymbol{\theta}}$ is

$$P_{\boldsymbol{\theta}}(G) = P_{\boldsymbol{\theta}}(G \mid |V| = n_G)P_{\boldsymbol{\theta}}(|V| = n_G) = P^*_{\boldsymbol{\theta}}(G) \cdot p_{n_G}, \tag{40}$$

with $P^*_{\boldsymbol{\theta}}(G)$ of the form

$$P^*_{\boldsymbol{\theta}}(G) = \frac{1}{c_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)} \exp\Big(\sum_{k\in\mathcal{I}} \theta_{1,i}\Gamma_k(G) + \sum_{k\in\mathcal{J}} \theta_{2,k} f_k(n)\Gamma_k(G)\Big), \tag{41}$$

with $\mathcal{J} \subseteq \mathcal{I}$, cardinality $q \leq p$ and $\theta_{2,k} \neq 0$ for all $k \in \mathcal{J}$. $\mathcal{J}$ is defined as the set of social mechanisms that are not a constant function of network size. Finally, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $c_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is a finite normalising constant,

$$c_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \sum_{G\in\mathcal{G}^n} \exp\Big(\sum_{k\in\mathcal{I}} \theta_{1,k}\Gamma_k(G)) + \sum_{k\in\mathcal{J}} \theta_{2,k} f_k(n)\Gamma_k(G))\Big).$$

In these expressions, f_i are strictly increasing functions on the number of agents and they account for the effect of network size on the linking process. $\boldsymbol{\theta}_1 = (\theta_{1,1}, .., \theta_{1,p}), \boldsymbol{\theta}_2 = (\theta_{2,1}, ..., \theta_{2,q})$ are parameters to be estimated and reflect the relative importance a network statistics has on explaining the network observed in reality.

One of the main contribution of *f*ERGM is to lay the foundations of statistical network models that do not assume that network size tends to infinity. This is done by showing that maximum likelihood estimator MLE is consistent on sample size (see SI) and by showing that the non-existence problem of MLE in ERGM (Handcock et al., 2003), and inherited by existing methodologies for analysing multiple networks, is solved when the sample size tends to infinity (see SI).

4.3 Empirical study

In this section, we address the problem of determining if two of our theoretical claims are consistent with empirical data: we study if the sign of $\theta_{2,i}$ and the functional form of the terms

f_i (as stated in the previous section) are consistent with what is found in real-world networks. Then, we show how to perform tests for constant reciprocity and transitivity using fERGM.

A total of 214 directed networks were collected from four different settings. First, 84 of them are friendship networks between students in middle and high school in the United States Harris et al. (2009); Moody (2001); Goodreau et al. (2009); Goel and Salganik (2010). We denote this dataset by 84-US. The size of the giant connected component of these networks ranges between 25 and $2,539$. Second, 75 social networks of rural villages in southern India (denoted 75-IN); the links are constructed based on the question: "In your free time, whose house do you visit?" Jackson et al. (2012); Banerjee et al. (2013). The ranges of network sizes is between 164 and 782 individuals. Third, 36 friendship networks (36-US) between students in high school classrooms in the US (denoted 36-US) (McFarland, 2001). Their size ranges between 15 and 30 students. Fourth, 19 networks (19-DU) defined by relations of emotional support among students in 19 classrooms in Dutch high schools Snijders and Baerveldt (2003); Baerveldt et al. (2008). The size of these networks range between 31 and 91 students.

In all the analyses performed, we assume homogeneous populations across networks within the same dataset and between agents within the same network. Although the first assumption is quite strong, it has been pointed out that it is implicitly assumed in results whenever data consists of multiple networks and a single network was analysed (Lazega and Snijders, 2015), and it is also assumed in previous multilevel network analysis. Relaxing the second is possible by including agents' attribute like gender. However, we do not do this, as doing so would blur our message: showing the existence of regularity that has to be considered when interpreting and generalising network analyses on a single network, and that should be taken into consideration when performing statistical analysis on multiple networks.

We apply the same ERGM to each of the networks in the four datasets (Hunter et al., 2008). First, we consider ERGM defined solely by the network statistics number of links l. Figure 8 shows a strong negative linear relation between θ_l and $\log$ of network size for all datasets (R^2 ranges between 0.48 and 0.95); therefore, we can conclude that $\theta_{2,l} < 0$ and $\log(n)$ is a good

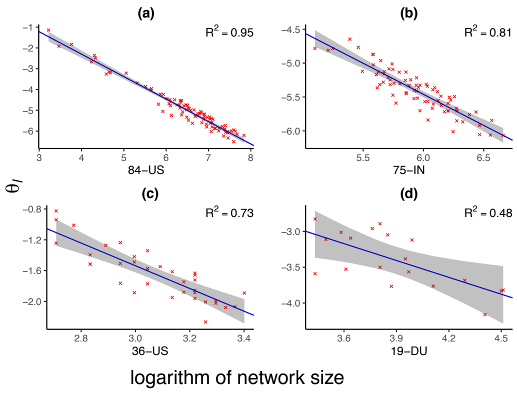


Figure 8: Plot estimated parameters θ_l against the $\log$ of the number of agents, n. For each network in each dataset, an ERG model was run, and the model was defined by the network statistic number of links l. All plots show that θ_l is negatively correlated with network size, and there is a linear relation between θ_l and $\log(n)$.

approximation for the function $f_l(n)$. These results are consistent with our theoretical claims of $\theta_{2,l} < 0$ and $f_l = \log(n)$. An exploratory two-step procedure suggests that for all datasets $\theta_{2,l}$ are negative.

Next, we incorporate the effect of reciprocity by adding the network statistics number of reciprocal pairs of links r to the previous class of ERGMs. We follow a similar procedure to the described for above. Figure 9 shows that $\log(n)$ is again a good approximation of $f_l(n)$ (R^2 ranges between 0.58 and 0.94) and $\theta_{2,l}$ is negative, which are consistent with both: the previous model and our theoretical claims. Figure 9 also shows a strong negative linear relation between θ_r and $\log$ of network size for all datasets (R^2 ranges between 0.14 and 0.47), i.e. $\theta_{2,r} > 0$. Further, for all the datasets the estimated $\theta_{2,r}$ are positive.

Another important social mechanism that might depend on network size is transitivity. First, it is important to mention that this model gives results consistent with the previous ones regarding $\theta_{2,l} < 0$ and the real function is well approximated by $\log(n)$. For all datasets, there is a linear relation between θ_r and $\log(n)$. For the datasets 84-US and 36-US the relationship is positive (R^2 equal to 0.29 and 0.52, respectively and $\theta_{2,r} > 0$) while for the other datasets we cannot make any statement. Figure 10 also suggests that there is a linear relation between θ_t and $\log(n)$. For the 84-US and 75-IN datasets, the relationship is positive (R^2 equal 0.44

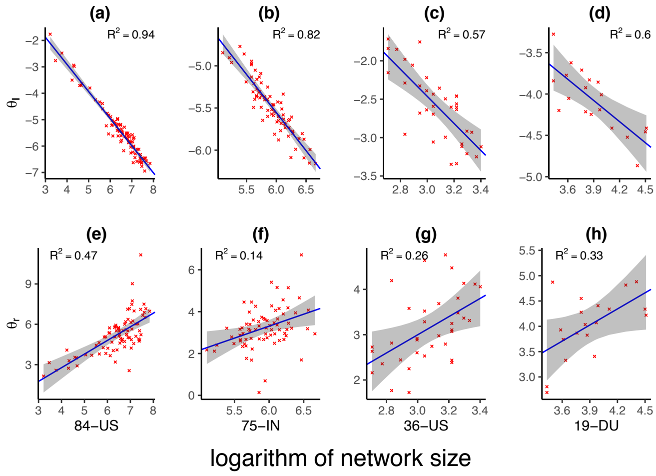


Figure 9: For each network in each dataset, an ERG model was run, and the model was defined by the network statistics number of links *l* and the number of reciprocal pairs *r*. In plots (a)-(d) we see that for all datasets, θ_l is negatively correlated with network size, and there is a linear relation between θ_l and $\log(n)$. In plots (e)-(h) we see that for all datasets, θ_r is positively correlated with network size. We see that there is a linear relation between θ_r and $\log(n)$.

and 0.69, respectively and $\theta_{2,r} > 0$) while for the other two datasets we cannot establish any positive nor negative relationship. For f_{m2p}, we see a linear relation between θ_{m2p} and $\log(n)$. However, only for the dataset 19-DU we can conclude that the relationship is positive (R^2 0.29 and θ_{m2p}).

4.3.1 Hypothesis testing

In this section, we demonstrate the power of *f*ERGM to perform hypothesis testing. We resort on two simple cases, by testing the hypothesis $\theta_{2,i} = 0$, and whether reciprocity and transitivity are constant on n. The construction of the following *f*ERGMs is done by assuming that f_i is equal to $\log(n)$ for all the interaction terms mentioned below.

The first *f*ERGM is defined by a single network statistics, the number of links and its interaction term with $\log(n)$. For all datasets, $\theta_{2,l}$ is negative ($p-$values < 0.05, see SI), which is consistent with our theoretical claims.

The second model extends the previous one by including the number of reciprocated pairs of links and its interaction term with $\log(n)$. We estimate the model for all datasets except for

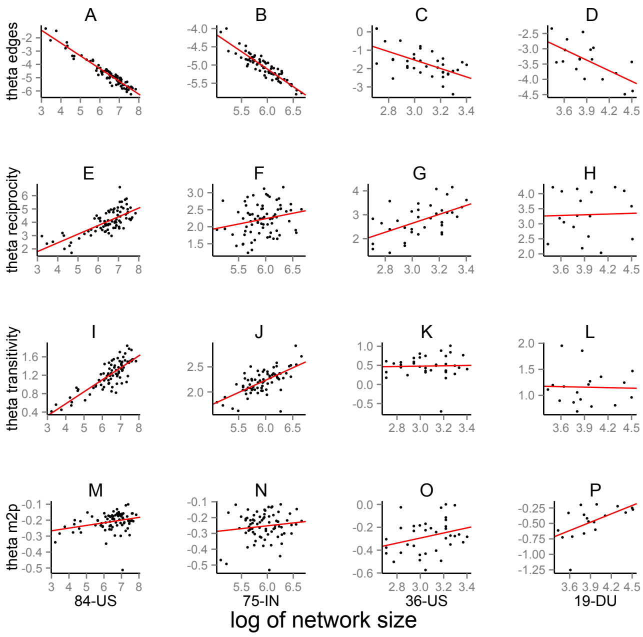


Figure 10: In plots (a)-(d) we see that for all datasets, θ_l is negatively correlated with network size, and there is a linear relation between θ_l and $\log(n)$. In plots (e)-(g) we see that for the three datasets, θ_r is positively correlated with network size and there is a linear relation between θ_r and $\log(n)$. However, this positive correlation is not clear for the dataset 19-DU. In plots (i)-(j) we see that for the two largest datasets, θ_t is positively correlated with network size and there is a linear relation between θ_r and $\log(n)$. However, this positive correlation is not clear for the datasets 36-US and 19-DU. In plots (m)-(p) we see that for all datasets, θ_{m2p} is positively correlated with network size, and there is a linear relation between θ_{m2p} and $\log(n)$.

the dataset 84-US, in which we encounter convergence problem (see SI). For the three other datasets, $\theta_{2,l}$ is negative ($p-$value < 0.05) and $\theta_{2,r}$ is positive ($p-$value < 0.05); the tests are performed by approximating the statistics by a normal distribution as in the previous model. These results are consistent with both: our theoretical arguments, and the empirical findings based on the exploratory two-step procedure.

Testing for constant reciprocity is done using a χ^2 test with one degree of freedom. For all datasets, we cannot reject the null hypothesis of constant reciprocity, see Table 4 Model 2.

Finally, we add two further network statistics to the second fERGM model: number of triangles and number of mixed two-paths and their interaction terms with the logarithm of the number of agents. In this way, we incorporate the tendency of agents to close a transitive relation. Similar to the previous model, we test for $\theta_{2,i} = 0$ with $i \in l, r, t, m2p$, constant reciprocity ($\theta_{2,r} = -\theta_{2,l}$) and constant transitivity ($\theta_{2,t} = -\theta_{2,l}$) using a χ^2 test with one degree of freedom for the three datasets 75-IN, 36-US and 19-DU, see Table 4 Model 3. As in our previous model, the estimated parameters for $\theta_{2,l}$ ($p-$value<0.05) are negative and for $\theta_{2,r}$ the values are positive for the three datasets 75-IN, 35-US and 19-US ($p-$value<0.05). For the datasets 75-IN and 19-DU, the estimated values of $\theta_{2,l}$ are positive ($p-$value<0.05), but negative for the dataset 36-US ($p-$value<0.05). For all three datasets, the estimated parameters for $\theta_{2,m2p}$ are positive ($p-$value<0.05). Constant reciprocity is only rejected for the dataset 75-IN ($p-$value<0.05), and it is also possible to reject the hypothesis of constant transitivity for the dataset 36-US ($p-$value<0.05).

4.4 Discussion

Hypothesis testing in networks has grown in importance during the last decades, and in this context, ERGM has become an important tool for social scientists. Notwithstanding, the ERGM popularity among practitioners, some problems persist, such as it remains largely unknown under which conditions are ERGM estimators consistent, and if they are, it is still insufficient to justify conclusions based on observed networks with a few dozen of agents. First, large sample properties used for hypothesis testing do not necessarily hold for small networks. Second, parameters are not a constant function of network size, and thus it is hard to compare results across studies, even if the subject is arguably the same.

To address the problems mentioned above, we have shown that under some mild conditions, $\log(n)$ is a good approximation for the dependency of basic network statistics (number of links, number of reciprocated links and number of transitive triangles) on network size. We have also introduced a new class of models, i.e. finite exponential random graph models fERGM which give consistent estimators on sample size. In this context, substituting assumptions on large networks by ones on large samples is a prerequisite for studying small networks and also to test for the effect the number of agents has on the formation of networks. Further, we demonstrated that the fERGM can be used to perform hypothesis testing, like verifying if social mechanisms are constant with the network size or not. Based on the $\log(n)$ approximation of the parameters for modeling reciprocity and transitivity, we have shown how fERGM can be used to test if these mechanisms are constant on the size of the network. Contrary to the established methodology to test if environment affects the formation of networks Lubbers (2003), the estimators constructed with the fERGM are shown to be consistent on the number of observed networks and their asymptotic distribution is known.

Several challenges lie ahead. First, further studies of the effects that group-level variables have on the formation of networks have to extend our analysis of the functional form of the parameters on n to other network statistics and other conditions. Second, it is clear that the models we studied (while complex enough to observe regularities between parameters and network size), are still too simple to control for all variables that have an effect on the linking process. For example, extending the model to control for agents' attributes is a prerequisite for a future studies testing hypothesis about network formation processes. Our proposed approach might also be seen as an alternative to the usage of complex network statistics introduced in models aimed to solve the non-existence problem of maximum likelihood estimators common in empirical studies (Snijders et al., 2006; Lubbers and Snijders, 2007).

Table 4: Test for constant reciprocity and constant transitivity using *f*ERGM model.

Model	Test	84-US	75-IN	36-IN	19-DU
Model 1	$\theta_{2,r} = -\theta_{2,l}$ (χ^2_1)		$3e-4$	0.026	0.002
Model 3	$\theta_{2,r} = -\theta_{2,l}$ (χ^2_1)		59.3^{***}	0.12	0.22
	$\theta_{2,t} = -\theta_{2,l}$ (χ^2_1)		$9e^{-4}$	18.82^{***}	1.13
Note:					$^{*}p$ <0.1; $^{**}p < 0.05$; $^{***}p < 0.01$

References

Baerveldt, C., Völker, B., and Van Rossem, R. (2008). Revisiting selection and influence: An inquiry into the friendship networks of high school students and their association with delinquency 1. *Canadian Journal of Criminology and Criminal Justice*, 50(5):559–587.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144):1236498.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Besag, J. (1975). Statistical analysis of non-lattice data. *The statistician*, pages 179–195.

Boda, Z. and Néray, B. (2015). Inter-ethnic friendship and negative ties in secondary school. *Social Networks*, 43:57–72.

Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical linear models: applications and data analysis methods.* Sage Publications, Inc.

Camacho Guardian, A. (2016). On evidence and social networks.

Camacho Guardian A., Tessone, C. and Algesheimer, R. (2017). Power law random graph models.

Chandrasekhar, A. G. and Jackson, M. O. (2014). Tractable and consistent random graph models. Technical report, National Bureau of Economic Research.

Collaboration, O. S. et al. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716.

Cranmer, S. J. and Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Political Analysis*, 19(1):66–86.

Cranmer, S. J., Desmarais, B. A., and Menninga, E. J. (2012). Complex dependencies in the alliance network. *Conflict Management and Peace Science*, 29(3):279–313.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., and Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest*, 2(1):1–30.

Ellwardt, L., Wittek, R., and Wielers, R. (2012). Talking about the boss: Effects of generalized and interpersonal trust on workplace gossip. *Group & organization management*, pages 521–549.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.

Goel, S. and Salganik, M. J. (2010). Assessing respondent-driven sampling. *Proceedings of the National Academy of Sciences*, 107(15):6743–6747.

Goodreau, S. M., Kitts, J. A., and Morris, M. (2009). Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. *Demography*, 46(1):103–125.

Guo, G. and Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology*, 26(1):441–462.

Handcock, M. S., Robins, G., Snijders, T., Moody, J., and Besag, J. (2003). Assessing degeneracy in statistical models of social networks. In *Journal of the American Statistical Association*.

Harris, K. M., Halpern, C. T., Whitsel, E., Hussey, J., Tabor, J., Entzel, P., and Udry, J. R. (2009). The national longitudinal study of adolescent to adult health: Research design. *www document, URL: http://www.cpc.unc.edu/projects/addhealth/design*.

Huitsing, G. and Veenstra, R. (2012). Bullying in classrooms: Participant roles from a social network perspective. *Aggressive Behavior*, 38(6):494–509.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3).

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008a). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3):54860.

Jackson, M. O., Rodriguez-Barraquer, T., and Tan, X. (2012). Social capital and social quilts: Network patterns of favor exchange. *The American Economic Review*, pages 1857–1897.

Jasny, L., Waggle, J., and Fisher, D. R. (2015). An empirical examination of echo chambers in us climate policy networks. *Nature Climate Change*.

Kim, J.-H., Holman, D. J., and Goodreau, S. M. (2015). Using social network methods to test for assortment of prosociality among korean high school students.

Krivitsky, P. N., Handcock, M. S., and Morris, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Statistical methodology*, 8(4):319–339.

Kruse, H., Smith, S., van Tubergen, F., and Maas, I. (2016). From neighbors to school friends? how adolescents' place of residence relates to same-ethnic school friendships. *Social Networks*, 44:130–142.

Lazega, E. and Snijders, T. A. (2015). Multilevel network analysis for the social sciences.

Lewis, K. (2013). The limits of racial prejudice. *Proceedings of the National Academy of Sciences*, 110(47):18814–18819.

Louch, H. (2000). Personal network integration: transitivity and homophily in strong-tie relations. *Social networks*, 22(1):45–64.

Lubbers, M. J. (2003). Group composition and network structure in school classes: a multilevel application of the p* model. *Social Networks*, 25(4):309–332.

Lubbers, M. J. and Snijders, T. A. (2007). A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social Networks*, 29(4):489–507.

Lubell, M., Scholz, J., Berardo, R., and Robins, G. (2012). Testing policy theory with statistical models of networks. *Policy Studies Journal*, 40(3):351–374.

Lusher, D., Koskinen, J., and Robins, G. (2012). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press.

McFarland, D. A. (2001). Student resistance: How the formal and informal organization of classrooms facilitate everyday forms of student defiance. *American Journal of Sociology*, 107(3):612–678.

McFarland, D. A., Moody, J., Diehl, D., Smith, J. A., and Thomas, R. J. (2014). Network ecology and adolescent social structure. *American sociological review*, page 0003122414554001.

Moody, J. (2001). Peer influence groups: identifying dense clusters in large networks. *Social Networks*, 23(4):261–283.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.

Oberwittler, D. (2007). The effects of neighbourhood poverty on adolescent problem behaviours: A multi-level analysis differentiated by gender and ethnicity. *Housing Studies*, 22(5):781–803.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2):173–191.

Schaefer, D. R. and Simpkins, S. D. (2014). Using social network analysis to clarify the role of obesity in selection of adolescent friends. *American journal of public health*, 104(7):1223–1229.

Schweinberger, M. and Handcock, M. S. (2015). Local dependence in random graph models: characterization, properties and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):647–676.

Snijders, T. A. and Baerveldt, C. (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of Mathematical Sociology*, 27(2-3):123–151.

Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153.

Song, H. (2015). Uncovering the structural underpinnings of political discussion networks: Evidence from an exponential random graph model. *Journal of Communication*, 65(1):146–169.

Valente, T. W., Fujimoto, K., Chou, C.-P., and Spruijt-Metz, D. (2009). Adolescent affiliations and adiposity: a social network analysis of friendships and obesity. *Journal of Adolescent Health*, 45(2):202–204.

Vaquera, E. and Kao, G. (2008). Do you like me as much as i like you? friendship reciprocity and its effects on school outcomes among adolescents. *Social Science Research*, 37(1):55–72.

Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393(6684):440–442.

Wu, B., Jiang, S., and Chen, H. (2015). The impact of individual attributes on knowledge diffusion in web forums. *Quality & Quantity*, 49(6):2221–2236.

5 Supplementary Materials (B)

5.1 Results

Our main results on this section are about statistical properties of *f*ERGM, and two-step procedures as the number of observed networks tend to infinity. First, we show that the MLE is consistent for *f*ERGM (Theorem 2), and that it converge in distribution to the normal distribution (Theorem 3). As a corollary of Theorem 2, we show that *f*ERGM solves the non-existence problem of MLE (Corollary 1). Next, we show that two-step procedures do not give consistent estimators (Theorem 5), and they do not solve the non-existence problem of MLE (Lemma 1).

Assumption 3. *Let $\mathcal{I}$ be a finite index of network structures.*

1. *The conditional probability of a variable e_{ij} on the remaining variables has the form,*

$$P(e_{ij} \mid \boldsymbol{e}_{-ij}) = c_{ij} \prod_{k \in \mathcal{I}} \prod_{\substack{e_{ij} \subset \boldsymbol{e}_S \\ \boldsymbol{e}_S \in S_k}} Q_S^k(\boldsymbol{e}_S), \tag{42}$$

   for some constant c_{ij}

2. *If $e_S \in S_k$*

$$Q_S^k(\boldsymbol{e}_S) = \begin{cases} c_k & \textit{if } G[\boldsymbol{e}_S] \sim H_k \\ 1 - c_k & \textit{otherwise} \end{cases} \tag{43}$$

   for some $c_k \in (0, 1)$.

Theorem 1. *If Assumption 3 is satisfied, then the join distribution is*

$$P_{\boldsymbol{\theta}}(\boldsymbol{e}) = \frac{1}{c(\boldsymbol{\theta})} \exp(\sum_{k \in \mathcal{I}} \theta_k \Gamma_k(\boldsymbol{e})) \tag{44}$$

where $\Gamma_k(\boldsymbol{e})$ is called a network statistics, and it is the number of subnetworks in $G[\boldsymbol{e}]$ isomorphic to H_k and $\theta_k = \log(\frac{c_k}{1-c_k})$.

We will assume that we have a collection of networks $\boldsymbol{G} = \{G_i\}_{i=1}^m$ drawn independently from a fixed *f*ERGM. We will assume that the *f*ERGM is defined by p network statistics, $(\Gamma_1, ..., \Gamma_p)$, and $q(\leq p)$ interaction terms between the network statistics and the number of agents in the networks $(f_1\Gamma_{j_1}, ..., f_q\Gamma_{j_q})$. We will denote the true parameter of the *f*ERGM by $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}$ denotes any vector in $\mathbb{R}^{p+q}$, $\boldsymbol{p}$ is the true vector of probabilities of observing a network of a particular size, and $\hat{\boldsymbol{p}}$ is the empirical probability vector for the collection $\boldsymbol{G}$.

Let us define $\boldsymbol{S}$ the vector with first entries equal to the network statistics, i.e. $S_i = \Gamma_i$ for all $i \leq p$; and the last entries equal to the product of the functions f_i with its respective network statistic, i.e. $S_{p+i} = f_i\Gamma_{j_i}$ for all $i \leq q$.

$$\boldsymbol{S} = (\Gamma_1, ..., \Gamma_p, f_1\Gamma_{j_1}, ..., f_q\Gamma_{j_q})$$

$\mathrm{E}_{\boldsymbol{\theta}\boldsymbol{p}}(S)$ and $\mathrm{Cov}_{\boldsymbol{\theta},\boldsymbol{p}}(S)$ denote the expectation and covariance matrix for the fERGM defined by the parameter $\boldsymbol{\theta}$ and probability vector $\boldsymbol{p}$, respectively. $\mathrm{E}_{\boldsymbol{\theta},\hat{\boldsymbol{p}}}(S)$ and $\mathrm{Cov}_{\boldsymbol{\theta},\hat{\boldsymbol{p}}}(S)$ denote the expectation and covariance matrix when the true probability vector is replaced by the empirical probability vector.

The ML estimator for a collection of m networks $\{G_i\}_{i=1}^m$ is defined by

$$\hat{\boldsymbol{\theta}}^m = \arg\max_{\boldsymbol{\theta}} \prod_{i=1}^m P_{\boldsymbol{\theta}}(G_i)\hat{p}_{n_{G_i}} \tag{45}$$

Where $\hat{p}_{n_{G_i}}$ equals the proportion of networks of size n_{G_i} in the sample.

Theorem 2. *If $\mathrm{E}(\mathrm{Cov}_{\boldsymbol{\theta_0},\boldsymbol{p}}(S|n))$ is strictly positive-definitive, then*

$$\hat{\boldsymbol{\theta}}^m \xrightarrow{p} \boldsymbol{\theta_0} \tag{46}$$

Corollary 1. *If the assumptions of Theorem 2 are satisfied, then the indicator function that is one when the MLE exists and zero otherwise converge to one in probability,*

Corollary 1 is an immediate consequence of Theorem 2.

Theorem 3. *If* $\mathrm{Cov}_{\boldsymbol{\theta}_0,\boldsymbol{p}}(S)$ *is strictly positive-definitive and there exists* N *such that* $p_n = 0$ *for all* $n \geq N$*, then*

$$I^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^m)(\hat{\boldsymbol{\theta}}^m - \boldsymbol{\theta}_0)^T \xrightarrow{D} \mathcal{N}(0, \mathbb{1}) \tag{47}$$

I *denotes the information matrix and* $\mathbb{1}$ *is the identity matrix of order* $p+q$*.*

The following theorem gives conditions to test if a social mechanism is part of the network formation process.

Theorem 4. *Let* $\boldsymbol{R}$ *be any vector in* $\mathbb{R}^{p+q}$*. If the assumption of Theorem 3 is satisfied; then*

$$(\boldsymbol{R} \cdot (\hat{\boldsymbol{\theta}}^m - \boldsymbol{\theta}_0))(\boldsymbol{R} I(\hat{\boldsymbol{\theta}}^m)\boldsymbol{R}^T)(\boldsymbol{R} \cdot (\hat{\boldsymbol{\theta}}^m - \boldsymbol{\theta}_0))^T \xrightarrow{D} \chi^2_1 \tag{48}$$

The above theorem can be used to test for constant reciprocity with a model including the network statistics *l*, r, *t* and *m2p* and under the assumption that $f_l(n) = f_r(n)$. This is done by taking as the null hypothesis $\theta_{2,r} = -\theta_{2,l}$, i.e. constant reciprocity; and by defining

$R = (0,0,0,0,1,1,0,0)$ and $\boldsymbol{\theta} = (\theta_{1,l}, \theta_{1,r}, \theta_{1,t}, \theta_{1,m2p}, \theta_{2,l}, \theta_{2,r}, \theta_{2,t}, \theta_{2,m2p})$

Similarly, we can test for constant transitivity by taking as the null hypothesis $\theta_{2,t} = -\theta_{2,l}$, i.e. constant transitivity.

Theorem 5. *Let* $\{G_i\}_{i=1}^{\infty}$ *be a sequence of independent exponential random graphs with* $G_i \sim P_{\theta_i}$*. Let us assume that* $P_{\boldsymbol{\theta}(i)}$ *are probability distributions defined as in Equation (44), with the same* p *network statistics but with different parameters. For each statistics* k*, we assume*

$\theta_k(i) = \theta_{1,k} + \theta_{2,k} f_k(n_i) + \epsilon_i$ *with* ϵ_i *independent and identically distributed,* ϵ*, with* $\mathrm{E}(\epsilon) = 0$*,* $\mathrm{Var}(\epsilon) = \sigma$*,* n_i *denotes the number of agents in the* i^{th} *network. If Assumption 4 is satisfied, then for an infinite number of networks, the MLE does not exist.*

Assumption 4. *One of the following propositions is satisfied:*

1. *An infinity number of networks are drawn from the same model.*

2. *The network size takes a finite number of values and* ϵ *is normally distributed.*

5.2 Proofs

The proof of Theorem 1 is in (Camacho Guardian, 2016).

The proof of Theorem 2 relies on the following proposition See (Newey and McFadden, 1994)[Theorem 2.7 page 2133]

Theorem 6. *If there is a function* $L(\boldsymbol{\theta})$ *such that (i)* $L(\boldsymbol{\theta})$ *is uniquely maximised at* $\boldsymbol{\theta}_0$*, (ii)* $\boldsymbol{\theta}_0$ *is an element of the interior of a convex set* Θ*, (iii)* $\hat{L}_m(\boldsymbol{\theta})$ *is concave; and (iv)* $\hat{L}_m(\boldsymbol{\theta}) \xrightarrow{p} L(\boldsymbol{\theta})$ *for all* $\boldsymbol{\theta} \in \Theta$*, then* $\hat{\boldsymbol{\theta}}_m$ *exists with probability approaching one and* $\hat{\boldsymbol{\theta}}_m \xrightarrow{p} \boldsymbol{\theta}_0$

Proof of Theorem 2

Our function L is the log likelihood function

$$L(\theta) = \mathrm{E}_{\boldsymbol{\theta}_0}(\log(p_{\boldsymbol{\theta}}(G)) \tag{49}$$

(i) Recall that $\log(t) \leq t - 1$ and it is strict when $t \neq 1$; and let μ be a probability measure such that the family of probability measures $(P_\theta)_{\theta\in\Theta}$ is dominated by μ.

$$L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}_0) = \int (\log(p_{\boldsymbol{\theta}}(G)) - \log(p_{\boldsymbol{\theta}_0}(G)))p_{\boldsymbol{\theta}_0}(G)d\mu(G) = \int log(\frac{p_{\boldsymbol{\theta}}(G)}{p_{\boldsymbol{\theta}_0}(G)})p_{\boldsymbol{\theta}_0}(G)d\mu(G) <$$

$$\int (p_{\boldsymbol{\theta}}(G) - p_{\boldsymbol{\theta}_0}(G))d\mu(G) = \int (p_{\boldsymbol{\theta}}(G) - p_{\boldsymbol{\theta}_0}(G))d\mu(G) = 1 - 1 = 0$$

Thus, $L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}_0) < 0$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

(ii) Follows by assuming that the parameter space is all $\mathbb{R}^{p+q}$, i.e. $\Theta = \mathbb{R}^{p+q}$

(iii) To show that $\hat{L}_m(\boldsymbol{\theta})$ is a concave function, it is sufficient to show that the Hessian of the function $\hat{L}_m(\boldsymbol{\theta})$ is equal to $\mathrm{E}_{\hat{\boldsymbol{p}}}(-\mathrm{Cov}_{\boldsymbol{\theta}}(\boldsymbol{S}|n))$.

$$\frac{\partial L_m(\boldsymbol{\theta})}{\partial \theta_{1,k}} = \frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial \theta_{1,k}} \log(p_{\boldsymbol{\theta}}(G_i)) = \frac{1}{m}\sum_{i=1}^{m}(\Gamma_k(G_i) - \mathrm{E}_{\boldsymbol{\theta}}(\Gamma_k \mid n = n_i)) = \bar{\Gamma}_k - \mathrm{E}_{\theta,\hat{\boldsymbol{p}}}(\Gamma_k) \tag{50}$$

$$\frac{\partial L_m(\boldsymbol{\theta})}{\partial \theta_{2,k}} = \frac{1}{m}\sum_{i=1}^{m}(f_k(n_i)\Gamma_k(G_i) - f_k(n_i)\mathrm{E}_{\theta}(\Gamma_k \mid n = n_i)) = \frac{1}{m}\sum_{i=1}^{m}(f_k(n_i)\Gamma_k(G_i)) - \mathrm{E}_{\boldsymbol{\theta},\hat{\boldsymbol{p}}}(f_k\Gamma_k) \tag{51}$$

Deriving Equations (50) and (51) gives,

$$\frac{\partial^2 L_m(\boldsymbol{\theta})}{\partial \theta_{1,k_1}\partial \theta_{1,k_2}} = \frac{1}{m}\sum_{i=1}^{m} -\mathrm{Cov}_{\boldsymbol{\theta}}(\Gamma_{k_1}, \Gamma_{k_2} \mid n = n_i) = \mathrm{E}_{\hat{\boldsymbol{p}}}(-\mathrm{Cov}_{\boldsymbol{\theta}}(\Gamma_{k_1}, \Gamma_{k_2}|n)) \tag{52}$$

$$\frac{\partial^2 L_m(\boldsymbol{\theta})}{\partial \theta_{1,k_1}\partial \theta_{2,k_2}} = \mathrm{E}_{\hat{\boldsymbol{p}}}(-\mathrm{Cov}_{\boldsymbol{\theta}}(\Gamma_{k_1}, f_{k_2}\Gamma_{k_2}|N)) \tag{53}$$

$$\frac{\partial^2 L_m(\boldsymbol{\theta})}{\partial\theta_{2,k_1}\partial\theta_{2,k_2}} = \mathrm{E}_{\hat{\boldsymbol{p}}}(-\mathrm{Cov}_{\boldsymbol{\theta}}(f_{k_1}\Gamma_{k_1}, f_{k_2}\Gamma_{k_2}|n)) \tag{54}$$

(iv) $\hat{L}_n(\boldsymbol{\theta}) \to L(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$ follows by the law of large numbers.

■

Proof of Theorem 3 The proof of Theorem 3 relies on the following proposition, see (Newey and McFadden, 1994)[Theorem 2.3 page 2146]

Theorem 7. *Suppose that $G_1, ..., G_m$ are i.i.d, the assumptions of Theorem 6 are satisfied, (i) $\boldsymbol{\theta}_0$ is an element of the interior of a convex set Θ, (ii) $p(G|\boldsymbol{\theta})$ is twice differentiable and $p(G|\boldsymbol{\theta}) > 0$ in a neighbourhood $\mathcal{B}$ of $\boldsymbol{\theta}_0$, (iii) $\int \sup_{\boldsymbol{\theta}\in\mathcal{B}} ||\nabla_{\boldsymbol{\theta}} p(G|\boldsymbol{\theta})||dp(G) < \infty$, $\int \sup_{\boldsymbol{\theta}\in\mathcal{B}} ||\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} p(G|\boldsymbol{\theta})||dp(G) < \infty$; (iv) $I = \mathrm{E}(\nabla_{\boldsymbol{\theta}} \log(p(G|\boldsymbol{\theta_0}))\nabla^t_{\boldsymbol{\theta}} \log(p(G|\boldsymbol{\theta}_0)))$ exists and is nonsingular, (v) $\mathrm{E}(\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \times \sup_{\boldsymbol{\theta}\in\mathcal{B}} ||p(G|\boldsymbol{\theta})||)$,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^m - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}(0, I) \tag{55}$$

Proof:

(i) It was proven in Theorem 2.

(ii) $p(G|\boldsymbol{\theta}) > 0$ follows from the fact that $\forall\boldsymbol{\theta} \in \mathbb{R}^p$ and $\forall G$

$$\exp\Big(\sum_{k\in\mathcal{I}} \theta_{1,i}\Gamma_k(G) + \sum_{k\in\mathcal{I}} \theta_{2,k} f_k(n)\Gamma_k(G)\Big) > 0 \tag{56}$$

Twice differentiability follows from the proof of Theorem 2.

(iii) Let $\mathcal{B}$ be a bounded interval of $\boldsymbol{\theta}_0$, and $\mathcal{C}$ a compact set containing $\mathcal{B}$. By definition $\nabla_{\boldsymbol{\theta}} p(G|\boldsymbol{\theta})$ is continuous on $\boldsymbol{\theta}$, and thus for a fixed network G, we have $\sup_{\boldsymbol{\theta}\in\mathcal{C}} ||\nabla_{\boldsymbol{\theta}} p(G|\boldsymbol{\theta})||$ is finite. In particular, $\sup_{\boldsymbol{\theta}\in\mathcal{B}} ||\nabla_{\boldsymbol{\theta}} p(G|\boldsymbol{\theta})||$ is finite.

Therefore, $\int \sup_{\boldsymbol{\theta}\in\mathcal{B}} ||\nabla_{\boldsymbol{\theta}} p(G|\boldsymbol{\theta})||dp(G)$ is the sum of finite numbers, and thus the integral exists.

Using similar arguments, we have that $\int \sup_{\boldsymbol{\theta}\in\mathcal{B}} ||\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} p(G|\boldsymbol{\theta})|| dp(G)$ is finite.

(iv) For the nonsingularity notice $E(\frac{\partial \log(p(G|\boldsymbol{\theta}_0))}{\partial \theta_{1,k_1}} \frac{\partial \log(p(G|\boldsymbol{\theta}_0))}{\partial \theta_{1,k_2}}) = \mathrm{E}(\mathrm{Cov}(\Gamma_{k_1}, \Gamma_{k_2}|n))$ by similar arguments to those used to derive Equation (53).

Thus, $\mathrm{E}(\nabla_{\boldsymbol{\theta}} \log(p(G|\boldsymbol{\theta}))\nabla^t_{\boldsymbol{\theta}} \log(p(G|\boldsymbol{\theta}))) = \mathrm{E}(\mathrm{Cov}(S|n))$, and the right side is by assumption nonsingular.

(v) It follows by the same arguments as in point (iii).

■

Proof of Theorem 4

Applying Slutsky's theorem to Formula (47), we have

$$(\hat{\boldsymbol{\theta}}^m - \boldsymbol{\theta}_0)^T R^T \xrightarrow{D} \mathcal{N}(0, RI(\hat{\boldsymbol{\theta}}^m)R^T) \tag{57}$$

Finally, recalling that the quadratic form of a normal distribution is a chi-square distribution, we get Formula (48).

■

Proof of Theorem 5

Before proving Theorem 5, let us introduce some notation. Let $C(\Gamma, n)$ the convex hull of the set $\{\boldsymbol{\Gamma}(G) \mid G \in \mathcal{G}^n\} \subset \mathbb{R}^p$, and C°, ∂C denote the interior and the boundary of C, respectively.

We will use the result shown in (Handcock et al., 2003) stating that if a graph G satisfies $\boldsymbol{\Gamma}(G) \in \partial C$, then $\hat{\boldsymbol{\theta}}(G)$ does not exist.

Lemma 1. *If assumption 4 holds, then the series in (58) diverges*

$$\lim_{m\to\infty} \sum_{i=1}^{m} P_{\boldsymbol{\theta}(i)}(G_i \in rbd(C)) \tag{58}$$

Proof of Theorem 5:

If the series in (58) diverges, then from the second Borell-Cantelli lemma it follows that when the sum of the probabilities of independent events diverges to infinity, then the probability that an infinite number of them occur is equal to one. In our case, an infinite number of times we observe a network with undefined MLE.

■

Proof of Lemma 1:

If point 1 in Assumption 4 is satisfied, then there exists $\{G_{n_i}\}_{i\in\mathbb{N}}$, a subsequence of observed networks drawn from the same model, such that for a $p > 0$

$$P_{\boldsymbol{\theta}}(G_{n_i} \in rbd(C)) = p \tag{59}$$

As a result, the series (58) is bounded by below by the sum of infinite p; which diverges.

For point 2, let us define

$$g(G,\delta) = \sum_{k\in\mathcal{I}} \theta_{1,k}\Gamma_k(G) + \sum_{k\in\mathcal{J}} \theta_{2,k} f_k(n_G)\Gamma_k(G) + \delta \tag{60}$$

and

$$f(n,\delta) = \frac{1}{c(n,\delta)} \sum_{G\in rbd(C(n,\Gamma))} \exp\left(g(G,\delta)\right) \tag{61}$$

with $c(n,\delta) = \sum_{G\in\mathcal{G}^n} \exp\left(g(G,\delta)\right)$.

By continuity of the function $f(n,\delta)$ on δ, if we take a compact set $D(\neq \emptyset)$, then

$$f(n,\delta_n) = \inf_{\delta\in D} f(n,\delta) > 0 \tag{62}$$

for some $\delta_n \in D$.

If the network size takes a finite number of values, then there exists an n_0 such that it is observed an infinite number of times. For an infinite number of random variables, we have that $g(G, \boldsymbol{\epsilon}) \in D$, and thus $P_{\boldsymbol{\theta}(n)}(G \in rbd(G))$ is bounded by $f(n, \delta)$. Thereby, the series (58) has as lower bound a sum of an infinite numbers of δ_n.

■

5.3 Estimation procedures

The computational complexity to estimate ERG models makes impossible in many cases to compute maximum likelihood estimators. Several alternatives to the MLE have been proposed MCMC-MLE (Monte Carlo Markov chain MLE) (Hunter and Handcock, 2006), and MPLE (maximum pseudo-likelihood estimation) (Besag, 1975; Strauss and Ikeda, 1990), with MCMC-MLE preferred over the others (Robins et al., 2007). For each network in all datasets, we estimate an ERGM using Monte Carlo Markov chain method (Robbins-Monro) (Hunter et al., 2008). First, we define the ERGM only by the network statistics number of links *l*; this is the simplest model, and it is equivalent to the Bernoulli random graph model (Lusher et al., 2012). In all cases, the model converge. For this model, it is not necessary to use MCMC-MLE, as computing MLE does not impose a significant computational burden.

Next, we define the ERGM by adding to the previous model the network statistics number of reciprocated links *r*. As this model is more complex than the previous one, some measures have to be taken for the MCMC-MLE to converge: we set the minimum number of steps in the MCMC to be 1 million.

The third model added to the previous model the network statistics number of transitive triangles and mixed two paths. Like the previous model, some measures have to be taken for the MCMC-MLE to converge: we set the minimum number of steps in the MCMC to be 1 million. This, though, is not enough as for some networks the MLE might not even exist (Handcock et al., 2003). As shown before, as the number of networks grows the probability that the MLE does not exist for some networks tends to one. This is, in fact, the case, as for one network in the 19-DU, the number of transitive triangles is equal to zero, and thus the MLE cannot exist. We remove this network from the analysis.

Our approach differs from the traditional estimation with ERGM in two aspects. First, modelling transitivity has been traditionally been done with transitive triangles (Frank and Strauss, 1986), but estimation problems have forced researchers to replace them with more complex statistics such as k-triangles, geometrically weighted edgewise shared partner- GWESP- (Snijders et al., 2006; Hunter, 2007; Goodreau et al., 2009). We do not use GWESP since transitive triangles have a simpler interpretation and the encountered estimation problems are solved with fERGM when a large sample of networks are collected.

The second difference has to do with the concept of near-degeneracy, which occurs when the estimated model places too much probability on particular graphs (Robins et al., 2007). This has been said to be caused by a poorly specified model (Goodreau et al., 2009; Li, 2015), and a solution has been to define a new model. To ensure that near-degeneracy is not a problem in empirical studies, some global properties of the estimated model are compared with the observed global properties (Lewis, 2013; Hunter et al., 2008a). However, our approach treats a single network as a single observation, and we cannot reject the idea that poor fitting is due to a small sample size. For instance, the removal of a network in the dataset 19-DU for the third model is a result of a small-sample-size problem. Thus, our analysis at this point is exploratory.

Determining $\theta_{2,i}$ were done with a two-step procedures, but contrary to the previous use of this procedures, we use it as an exploratory tool and not for making statistical inference (Lubbers, 2003). After estimating the model for each network and each dataset as described above; we perform a simple linear regression for each dataset, each model and each parameter in the model,

$$\theta_i = \theta_{1,i} + \theta_{2,i} \log(n) + \epsilon_i \tag{63}$$

with i being the network statistics under consideration and n the number of agents in the networks. The results are presented in Tables 5-11. Our results suggest that the estimated parameters are a linear function of $\log(n)$. For the class of ERGM defined by the network statistics number of links and number of reciprocate pair of links, Figure 11 does not suggest that a transformation is needed for the dataset 84-US and $\hat{\theta}_r$, but that there exist some outliers and heteroscedasticity. This is to be expected since there is a positive probability that the absolute value of the estimator is infinite.

Before the empirical analysis with *f*ERGM, we develop Monte Carlo Markov chain procedure to approximate the MLE for *f*ERGM which relied on the R package *ergm* (Hunter et al., 2008). Consistent with the previous three classes of ERGMs, we define three different *f*ERGMs. One that only incorporates the number of links and assumes $f_l(n) = \log(n)$, the second adds the network statistics pair of links and assumes $f_r(n) = \log(n)$, and the last one adds the network statistics number of transitive triangles and mixed two paths and assumes $f_t(n) = f_{m2p}(n) = \log(n)$. We test for $\theta_{2,i} = 0$ with a Z-test (Theorem 3), and we test for constant reciprocity and transitivity with a χ^2 test with one degree of freedom (Theorem 4). The results for the first model are presented in Table 12, the results of model 2 are presented in Table 13, and the results of model 3 are presented in Table 14.

Table 5: The parameter θ_l as function of $log(n)$ for the class of ERGM defined by the network statistic number of links *l*. Results obtained using two-step procedure.

	Datasets:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,l}$	-1.080	-0.917	-1.475	-0.784
	(0.026)	(0.052)	(0.152)	(0.198)
$\theta_{1,l}$	2.016	0.039	2.887	-0.348
	(0.169)	(0.311)	(0.469)	(0.775)
Observations	84	75	36	19
R^2	0.955	0.808	0.733	0.479

Table 6: The parameter θ_l as a function of $log(n)$ for the class of ERGM defined by the network statistics number of links *l* and number of reciprocal pair of links r. Results obtained using two-step procedure.

	Datasets:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,l}$	-1.026	-0.903	-1.763	-0.853
	(0.028)	(0.050)	(0.263)	(0.169)
$\theta_{1,l}$	1.195	-0.138	2.827	-0.751
	(0.179)	(0.295)	(0.809)	(0.660)
Observations	84	75	36	19
R^2	0.944	0.819	0.569	0.601

Table 7: The parameter θ_r as a function of $log(n)$ for the class of ERGM defined by the network statistics number of links *l* and number of reciprocal pair of links r. Results obtained using two-step procedure.

	Datasets:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,r}$	1.014	1.143	2.005	1.060
	(0.119)	(0.333)	$(0.587$	(0.365)
$\theta_{1,r}$	-1.287	-3.553	-3.017	-0.109
	(0.771)	(1.976)	(1.806)	(1.426)
Observations	84	75	36	19
R^2	0.472	0.139	0.255	0.332

Table 8: The parameter θ_l as a function of $log(n)$ for the class of ERGM defined by the network statistics number of links *l*, the number of reciprocal pair of links r, the number of transitive triangles *t* and the number of mixed two paths *m2p*. Results obtained using two-step procedure.

	Dependent variable:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,l}$	-0.962	-0.961	-2.301	-1.157
$\theta_{1,l}$	1.452	0.651	5.366	1.141
	(0.191)	(0.292)	(1.669)	(1.362)
Observations	84	75	36	18
R^2	0.929	0.840	0.346	0.411

Table 9: The parameter θ_r as a function of $log(n)$ for the class of ERGM defined by the network statistics number of links *l*, the number of reciprocal pair of links r, the number of transitive triangles *t* and the number of mixed two paths *m2p*. Results obtained using two-step procedure.

	Dependent variable:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,r}$	0.653^{***}	0.316^{**}	1.871^{***}	0.078
	(0.069)	(0.158)	(0.500)	(0.570)
$\theta_{1,r}$	-0.159	0.343	-2.968^{*}	2.993
	(0.449)	(0.938)	(1.539)	(2.242)
Observations	84	75	36	18
R^2	0.522	0.052	0.291	0.001
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 10: The parameter θ_t as a function of $log(n)$ for the class of ERGM defined by the network statistics number of links *l*, the number of reciprocal pair of links r, the number of transitive triangles *t* and the number of mixed two paths *m2p*. Results obtained using two-step procedure.

	Dependent variable:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,t}$	0.261	0.491	0.046	-0.033
	(0.019)	(0.064)	(0.251)	(0.260)
$\theta_{1,t}$	-0.462	-0.714	0.342	1.285
	(0.125)	(0.383)	(0.773)	(1.022)
Observations	84	75	36	18
R^2	0.692	0.443	0.001	0.001

Table 11: The parameter θ_{m2p} as a function of $log(n)$ for the class of ERGM defined by the network statistics number of links *l*, the number of reciprocal pair of links r, the number of transitive triangles *t* and the number of mixed two paths *m2p*. Results obtained using two-step procedure.

	Dependent variable:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,m2p}$	0.017	0.036	0.222	0.450
	(0.006)	(0.031)	(0.122)	(0.163)
$\theta_{1,m2p}$	-0.317	-0.469	-0.960	-2.235
	(0.040)	(0.184)	(0.375)	(0.643)
Observations	84	75	36	18
R^2	0.083	0.018	0.089	0.322

Table 12: *f*ERGM model defined by the network statistic number of links and its interaction term with the log of the number of agents in the network.

	Datasets:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,l}$	-1.08^{***}	-0.92^{***}	-1.48^{***}	-0.82^{***}
	(0.005)	(0.03)	(0.37)	(0.1)
$\theta_{1,l}$	2.17^{***}	0.07	2.9	-0.18
	(0.18)	(0.71)	(3.15)	(1.24)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 13: *f*ERGM model defined by the network statistics number of links and number of pair of links and their interaction term with the log of the number of agents in the network.

	Datasets:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,l}$		$-.90^{***}$	-1.80^{***}	-0.81^{***}
		(0.03)	(0.11)	(0.11)
$\theta_{1,l}$		-0.18	2.95^{***}	-0.91
		(0.7)	(0.7)	(1.36)
$\theta_{2,r}$		0.93^{***}	2.06^{***}	0.75^{**}
		(0.12)	(0.35)	(0.33)
$\theta_{1,r}$		-2.18	-3.24	1.15
		(3.51)	(2.30)	(4.14)
$\theta_{2,r} = -\theta_{2,l}$ (χ^2_1)		$3e-4$	0.026	0.002
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 14: fERGM model defined by the network statistics number of links, the number of pair of links, the number of transitive triangles and the number of mixed two paths and their interaction term with the $\log$ of the number of agents in the network.

	Datasets:			
	84-US	75-IN	36-US	19-DU
	(1)	(2)	(3)	(4)
$\theta_{2,l}$		-0.99^{***}	-2.16^{***}	-1.50^{***}
		(0.007)	(0.52)	(0.14)
$\theta_{1,l}$		0.73^{***}	5.64	2.34
		(0.22)	(4.13)	(1.98)
$\theta_{2,r}$		0.36^{***}	1.94^{**}	1.00^{**}
		(0.03)	(0.88)	(0.42)
$\theta_{1,r}$		0.55	-3.02	-0.34
		(0.74)	()	(6.03)
$\theta_{2,t}$		0.50^{***}	-0.28^{***}	0.08^{***}
		(0.0001)	(0.06)	(0.01)
$\theta_{1,t}$		-0.9^{***}	1.26^{**}	0.48^{**}
		(0.003)	(0.51)	(0.19)
$\theta_{2,m2p}$		0.05^{***}	0.52^{***}	0.21^{***}
		$(9.9e-5)$	(0.07)	(0.01)
$\theta_{1,m2p}$		-0.47^{***}	-1.92^{***}	-1.04^{***}
		(0.002)	(0.62)	(0.22)
$\theta_{2,r} = -\theta_{2,l}$ (χ^2_1)		59.3^{***}	0.12	0.22
$\theta_{2,t} = -\theta_{2,l}$ (χ^2_1)		$9e^{-4}$	18.82^{***}	1.13
Note:				*p<0.1; **p<0.05; ***p<0.01

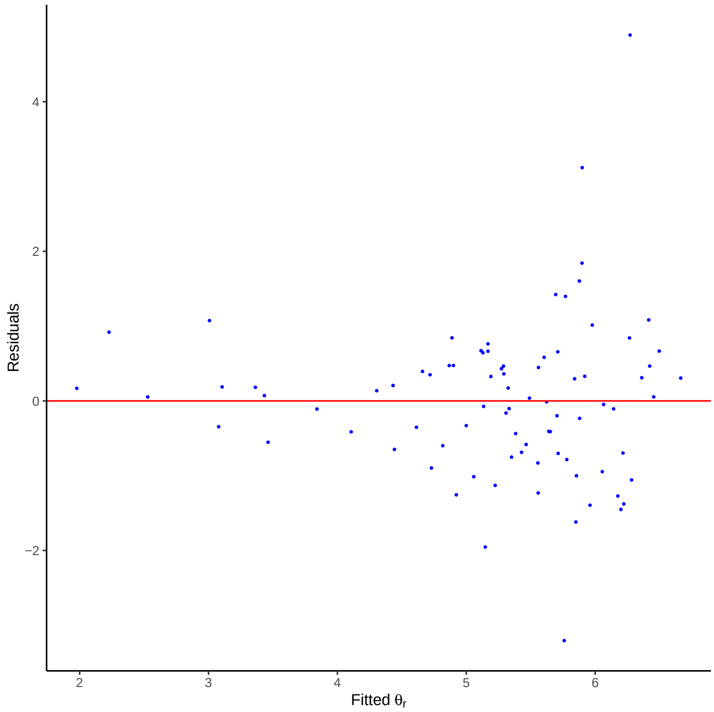


Figure 11: The estimated the residuals are plotted against the estimated θ_r for the model $\theta_r = \alpha + \beta \log(n) + error$.

6 Exponential random graph model with community structure (C-ERGM)

Abstract

In this paper, we review the well-known network model exponential random graph model (ERGM). First, we present a subclass of ERGM with a community structure, termed C-ERGM, and we give sufficient conditions on the community structure for having valid statistical inference for large networks, with "large network" being in terms of the number of communities in the network. Next, we show different axioms at the microscopic level used to construct subclasses of ERGM, e.g. Markov random graphs, partial conditional dependence, potential games. Then, we review existing issue when performing statistical inference with a single network. Contrary to the common view among ERGM practitioners that estimation problems suggest misspecified models, we emphasise that estimation problems are also caused by a small sample size problem, which is the result of observing networks with a single community. Then, we show shortcomings of one of the most popular approaches for analysing multiple networks. These shortcomings are the outcome of treating each network as consisting of several observations, irrespectively of the number of communities of the network. Finally, we discuss how the assumptions of a new method for analysing large networks implies that the formation of relations in a network can be considered as several independent processes.

6.1 Introduction

Social network science has undergone a period of transformations marked by an ascendancy of empirical work (Wasserman and Faust, 1994). At the root of this movement is the growing number of statistical models for analysing networks (Goldenberg et al., 2010). A goal of these network models is to provide statistical evidence based on observational or experimental data.

Examples statistical frameworks used to construct scientific evidence in social network studies are significance test (ST) (Fisher, 1922, 1935), hypothesis testing (HT) (Neyman and Pearson, 1992) and Null Hypothesis Significance Testing (NHST) (Gigerenzer, 2004). These statistical frameworks have been used to study the role network structures have in the formation of advice networks (Siciliano, 2015), determine the patterns and local structures in networks

Author Statement: This work is a single-author paper.

affected by organisational crisis (Uddin et al., 2013), study the network structures that are important for successful group outcomes (Lusher et al., 2013), explore knowledge diffusion in web forums (Wu et al., 2015), understand how states adjust their alliance portfolios (Cranmer et al., 2012) and testing whether structural balance is present or not in animal populations (Ilany et al., 2013).

In this paper, we review the network model exponential random graph model (ERGM). The importance of ERGM for the general public leans on at least two aspects. First, it has been used for studying social problems such as bullying (Huitsing and Veenstra, 2012), ethnic segregations (Lewis, 2013; Kruse et al., 2016), effects of overweight in friendship (Valente et al., 2009). Second, by studying how ERGM is used to construct scientific evidence, we contribute to the debate of false finding in Science (Collaboration et al., 2015; Head et al., 2015).

In Section 6.2, we provide a framework in which for a subclass of ERGM, termed C-ERGM, it is possible to justify statistical inference by assuming that agents are grouped in disjoint and independent communities and by observing network with a large number of communities. This contrast with the believe that inference is valid when the number of agents in a network is large (Hunter and Handcock, 2006). In Section 6.4, we review subclasses of ERGM defined by assumptions at the microscopic level, and in Section 6.5 we present necessary assumptions at the data-generating process for valid statistical inference.

In Section 6.6, we discuss problems emerging when making statistical inferences with a single network. First, we review the non-existence problem of the MLE and the convergence problem of Markov chain Monte Carlo methods for approximating MLE. Then, we present a simple example showing that letting the number of agents tend to infinity while fixing the number of communities to one is not sufficient for having valid statistical inference with network data. Our example is motivated by the mathematical results of Ellis and Newman (1978) which shows that in the Ising Courier Pott model, the central limit theorem may not hold when there is strong correlation between variables. Contrary to previous discussion affirming that estimation problems suggest misspecified models, in Section 6.6, we emphasize that estimation problems are also the result of small sample size problems, e.g. the number of communities in a network is one.

In Section 6.7, we show that the concept of communities in C-ERGM shares similarities to the concept of groups in Multilevel network models. These similarities are used to analyse Multilevel network models based on ERGM. Our main claim, here, is that each single small network should be treated as single observation and by consequence it is not possible to perform independent statistical analysis on them, as it is done in the popular two-step procedure for multilevel network analysis (Lubbers, 2003). Next, we compare the C-ERGM with the previous framework of local dependence condition introduced by Schweinberger and Handcock (2015). The local dependence condition searches for conditions that justify the central limit theorem for the network statistics and provides a framework for the construction of p-values. As C-ERGM, the local dependence condition groups the agents in disjoint and independent communities, so that a large network could be treated as several small networks.

Our goal in this review is to provide the reader with a critical view of the well-established network model ERGM. The critical position taken in this review is aimed to raise awareness among practitioners of limitations of network studies based on the exponential random graph model.

6.1.1 Statistical analysis on a single network and the p-value

In empirical work, conclusions drawn from data are uncertain, and therefore probability theory naturally arises as a tool to cope with uncertainty. In the statistical analysis of data, a research problem is to test a statement derived from theory; this statement is called the null hypothesis. The fist step in addressing this problem is to collect data and to summarise it with a test statistics. The (asymptotic) distribution of the test statistics has to be known under the null hypothesis and other auxiliary hypotheses (Cox, 1982). Next, a p-value is computed, which is, informally, defined as the probability of observing a value of the test statistic that is at least as extreme as the observed value assuming the null hypothesis, as well as other auxiliary assumptions, are true. If the auxiliary assumptions hold, the smaller the p-value, the greater the incompatibility of the data with the null hypothesis (Wasserstein and Lazar, 2016). Thus, when the p-value is small (i.e. when the test statistics is in the extreme of its probability distributions), we say there is evidence against the null hypothesis.

6.1.2 Limitation of statistical inference to construct scientific evidence

A major issue in empirical studies is that statistical evidence may be undermined by the implausibility of an auxiliary hypothesis (Smith, 2002). The nature of the auxiliary assumptions can be anything not contained in the theory such as assumptions on the data generating process (Freedman, 1987). Thus, when statistical analyses show small p-values, there are three possible explanations: (1) the null hypothesis is false, (2) the null hypothesis is true, but some auxiliary assumptions are false, or (3) a rare event has occurred. In network studies, statistical evidence based on regression models have been challenged by the independence assumption of these models (Lorenz and Hall, 2013; O'Malley, 2013). As a consequence, statistical network models are characterised by relaxing the independence assumption of regression models.

During the last years, there has been a growing interest in the network model exponential random graph model (ERGM), and its popularity is mostly due to the removal of the independence from the model assumptions. Although ERGM is intended to model the dependence in

When justifying hypothesis testing with finite-sample properties of OLS, we assume that the error is normal distributed. On the other hand, hypothesis testing with large sample theory does not require to assume that the error is normal distributed but that the sample is large. In experimental economics, objections on conclusions are often on the plausibility of an auxiliary assumption, e.g. payoffs are adequate to motivate subjects, subjects understood the instructions of the experiment (Smith, 2002; Ricciuti, 2008).

Most of the controversies surrounding statistical evidence is due to the auxiliary assumptions used for constructing a mathematical model. For instance, decisions based on statistical evidence in courtrooms have shown the adverse social effects of statistical evidence when models lean on unreasonable auxiliary assumptions (Meester et al., 2006).

An example of the controversy of the independence assumption is illustrated in (Lubell et al., 2012) "Policy theory is steeped in the empirical tradition of using descriptive statistics to describe data, and then moving to multivariate models that link dependent to independent variables, and relying on strong assumptions about individual units of analysis and the behavior of error terms. [...] For example, if one actor's performance is affected by its network relationships, then observations of actor performance are not independent, and general linear models that assume independence of observations may produce misleading estimates of the impact of an actor's number of network relationships ".

network data, there is little research on assumptions required to construct test statistics with known (asymptotic) distributions. Hence, it is custom to assume that the test statistics are approximately normally distributed when the sample is large as a result of the central limit theorem (Hunter and Handcock, 2006). For instance, as a rule, null hypothesis are rejected when an estimate (in absolute value) is greater than 2 times the standard error (Lusher et al., 2012, 2013). Although the central limit theorem is valid under general conditions of weak-independence (Hayashi, 2000), it does not hold when variables are strongly correlated (Ellis and Newman, 1978). Thus, weighing scientific evidence based on network data is only possible after understanding the auxiliary assumptions needed for the central limit theorem to hold.

6.2 Network Model

6.2.1 Definition and notation

We consider an infinite numerable set of agents $\mathbb{B}$ and a set of relational variables $\boldsymbol{e} = (e_{ij})_{ij \in \mathcal{F}}$ with $\mathcal{F} = \{ij \mid i, j \in \mathbb{B} \text{ with } i \neq j\}$. Throughout time, agents create and sever relations with each other, and the relations of an agent i are described by the set of variables $\{e_{ij} \mid j \in \mathbb{B} \text{ and } j \neq i\}$ with e_{ij} equal 1 if there is a relation (link) from agent i to j and zero otherwise. The set of agents and relational variables often represent biological, social, financial or communication networks.

As agents are generally found cluster in communities sharing common characteristics (Fortunato, 2010); our first assumption assigns each agent to a community.

Assumption 5. *There is a partition U of the agents into finite disjoint communities.*

$$\mathbb{B} = \bigcup_{l=1}^{\infty} U_l$$

The number of agents in a community is bounded by a constant B ($|U_l| < B$).

For several networks, we have reasonable knowledge about its community structure. For instance, communities might represents students in classrooms, schools, individuals in villages, towns, etc. However, defining communities is not always clear, and in some cases, communities are algorithmically defined without a precise *a priori* definition (Fortunato, 2010). In the following, it is assumed that the communities are well-defined and known.

In the following, we introduce a dependency structure between the relational variables to depict how agents update their relations under the restriction that only certain relations are conceivable. Specifically, for S a finite subset of $\mathcal{F}$ and any pair $ij \in S$, we consider the updating process of the variable e_{ij} when agent i exclusively considers the current value of the variables $\boldsymbol{e}_{S\setminus ij}$. For $\boldsymbol{e}_S$, the set of agents in a pair in S will be denoted by $V(\boldsymbol{e}_S)$; and a realisation of the variable is denoted by $\bar{\boldsymbol{e}}_S$.

The realisation $\bar{\boldsymbol{e}}_S$ is called an *hyponetwork* with agents $V(\boldsymbol{e}_S)$, see Figure 12 **(a)**. We call the hyponetwork $\bar{\boldsymbol{e}}_S$ a *network*, when for any two agents i, j in $V(\bar{\boldsymbol{e}}_S)$, we have that 1) $ij \in S$ or $ji \in S$, and 2) $ij, ji \in S$, implies that $i_2j_2, j_2i_2 \in S$ for all pair of $i_2, j_2 \in V(\bar{\boldsymbol{e}}_S)$. Point 1) affirms that between two agents there is always a decision to create a relation or not; while point 2) affirms that if for a pair of agent it is possible to have reciprocate relation, then it is possible for any other pair of agents to have a reciprocate relationship. In the case that $\bar{\boldsymbol{e}}_{S'}$ and $\tilde{\boldsymbol{e}}_S$ are (hypo)networks with $\boldsymbol{e}_{S'} \subseteq \boldsymbol{e}_S$ and $\tilde{\boldsymbol{e}}_S\big|_{\boldsymbol{e}_{S'}} = \bar{\boldsymbol{e}}_{S'}$, we say that $\bar{\boldsymbol{e}}_{S'}$ is a *sub(hypo)network* of $\tilde{\boldsymbol{e}}_S$. For a subset of agents $V' \subseteq V(\bar{\boldsymbol{e}}_S)$, the induced hyponetwork of V' is defined by the realisation of the variables $\bar{\boldsymbol{e}}_{S'}$ with $\boldsymbol{e}_{S'}$ being the maximal subset of variables satisfying $V' = V(\boldsymbol{e}_{S'})$. In Figure 12 **(a)**, the induced graph for the agents 3, 6 and 8 is described by the realisation

Relaxing the assumption that the size of the communities are finite is presented in Subsection 6.3.1.

$(e_{3,6}, e_{6,8}) = (1, 1)$, and the induced hyponetwork for the agents 3, 4 and 6 is described by the realisation $(e_{3,4}, e_{3,6}, e_{4,6}) = (1, 1, 0)$

A network $\bar{\boldsymbol{e}}_{S'}$ is a *directed network*, when $ij \in S$ and $ji \in S$, otherwise it is an undirected network. In the network literature, a similar concept to hyponetwork is "configuration". A configuration is defined by a set of agents (nodes) and some relations (links) (Robins et al., 2007), but a configuration does not distinguish if the absence of a relation between agents is because it is not possible or because agents decided not form a relationship.

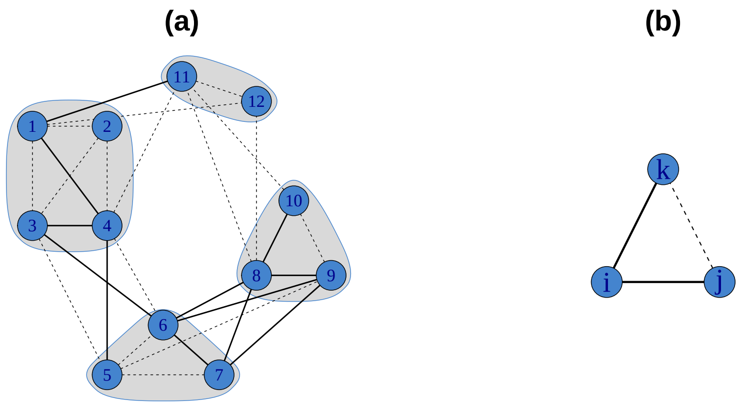


Figure 12: **(a)** illustrates an hyponetwork with 12 agents. The solid lines indicate the existence of relations between agents while the dot lines represent absence of relations between the agents. The non-existence of a line between agents implies that there is no relational variable defined between them. The induced hyponetworks for the agents 1, 3 and 4 is isomorphic to the hyponetwork in **(b)**. However, the induced hyponetwork for the agents 3, 6 and 8 is not isomorphic to the hyponetwork in **(b)**.

Two hyponetworks $\bar{\boldsymbol{e}}_{S'}$ and $\tilde{\boldsymbol{e}}_{S''}$ are said to be isomorphic ($\bar{\boldsymbol{e}}_{S'} \sim \tilde{\boldsymbol{e}}_{S''}$) if there is a bijection f between the agents that preserves relations and non-relations, i.e. $f : V(\bar{\boldsymbol{e}}_{S'}) \to V(\tilde{\boldsymbol{e}}_{S''})$ such that 1) for all $ij \in S'$, $\bar{e}_{ij}$ equals $\tilde{e}_{f(i)f(j)}$, and 2) for all $ij \in S''$, $\bar{e}_{f^{-1}(i)f^{-1}(j)}$ equals $\tilde{e}_{ij}$. Notice that by definition the induced hyponetwork defined by $\boldsymbol{e}_{S'} = \{e_{ij}, e_{ik}, e_{jk}\}$ and $(e_{ij}, e_{ik}, e_{jk}) = (1, 1, 0)$ is not isomorphic to the hyponetwork defined by $\boldsymbol{e}_{S''} = \{e_{ij}, e_{ik}\}$ and $(e_{ij}, e_{ik}) = (1, 1)$, see Figure 12 **(b)**. In the second case, we do not allow agent j to form a link to k while in the first case, agent j is allowed to choose to form a link to i and the choice is not to. Alternatively, we can say that the variables e_{ij} and e_{ik} are update without considering the value of the variable e_{jk}. In the Supplementary Materials, the definition of isomorphic hyponetworks is extended to include agents' attributes.

We call an isomorphism class of hyponetworks, $[H]$, a *structure*, and we said that an hyponetwork $\bar{\boldsymbol{e}}_{S'}$ posses structure H if $\bar{\boldsymbol{e}}_{S'} \sim H$, where H is a representative element in $[H]$. We said that agent $i \in V(\bar{\boldsymbol{e}}_{S''})$ is in a set of relations with structure H. For the sake of simplicity, when talking about a structure, we will often talk about a representative element.

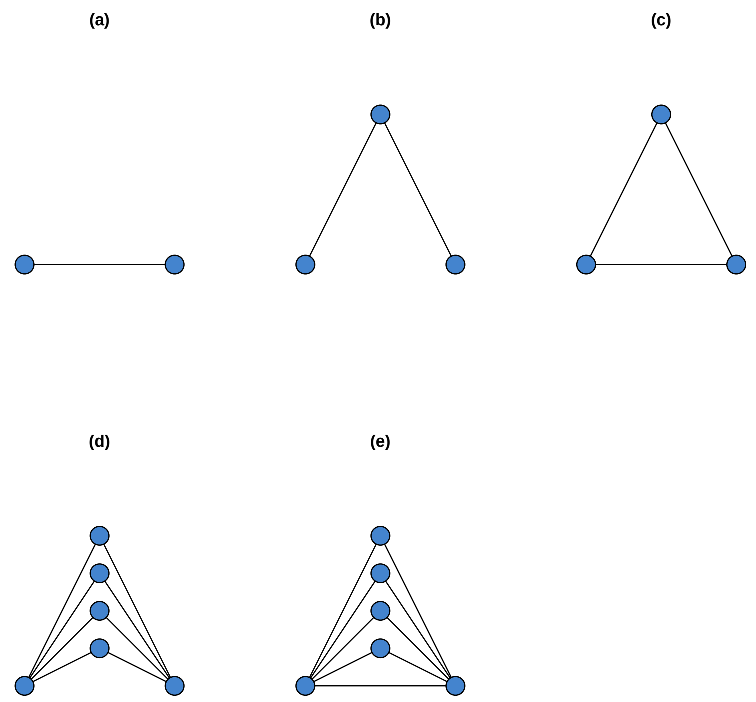


Figure 13: **(a)** illustrates a relation, **(b)** a mixed two path, **(c)** a triangle, **(d)** 4-two-stars and **(e)** illustrates 4-triangles. The absence of a line between two nodes indicates that the value of its associated relation variable is not considered or that there is no variable defined between the nodes.

6.2.2 Exponential random graph model with community structure

Given a set of relational variables $\boldsymbol{e}_S$, an agent i often prefers to update the relation $e_{ij} \in \boldsymbol{e}_S$ in such way that it is in an hyponetwork with a specific structure. For instance, when agent i decides to update the variable e_{ij} without considering the existence of other relations, agent i may prefer to update the variable such that it is in a relation with j, in which case he prefers to be in a hyponetwork with the structure defined by the two agents and e_{ij} being equal to one. Another example occurs when agent i updates the variable e_{ij} while considering solely the value of the variable e_{ji}. In this case, agent i may prefer to be in a hyponetwork with structure $e_{ij} = e_{ji} = 1$, e.g. agent prefers to reciprocate a relationship.

Next, we assume that the probability to update a relation conditioned on the others can be decomposed by a class of likelihoods (tendencies) of agents to belong to hyponetworks with certain structures. We denote the class of these structures by $\{H_k\}_{k=1}^{p}$, and we assume without loss of generality that H_1 is the structure with two agents i, j and the variable $e_{ij} = 1$, see Figure 13 (a). Other examples of structures are pair of reciprocate relations, mixed two path and transitive triad, see Figure 13 (b)-(d).

Let S_k denotes the set of subsets of relational variables such that if $\boldsymbol{e}_S \in S_k$, then there exists a realisation $\bar{\boldsymbol{e}}_S$ such that it possesses structure H_k with $k \in \{1, .., p\}$. S_k^l is the subset of S_k with $\boldsymbol{e}_S \in S_k^l$ if all agents in $V(\boldsymbol{e}_S)$ belong to the same community U_l, i.e. $V(\boldsymbol{e}_S) \subseteq U_l$.

Assumption 6. *For S a finite subset of $\mathcal{F}$ and for each $ij \in S$, we define the conditional probability distribution of a relational variable e_{ij} on the remaining relational variables as follows. If agent i and j belong to the same community U_l, then*

$$P^S(e_{ij} \mid \boldsymbol{e}_{S\setminus ij}) = c_{ij} Q_1(e_{ij}) \prod_{k=2}^{p} \prod_{\substack{e_{ij} \subseteq \boldsymbol{e}_{S'} \\ \boldsymbol{e}_{S'} \in S_k^l}} Q_{S'}^{(k,1)}(\boldsymbol{e}_{S'}), \tag{64}$$

with c_{ij} a constant.

If agent i and j belong to communities U_l, $U_{l'}$, respectively with $l \neq l'$, then

$$P^S(e_{ij} \mid \boldsymbol{e}_{S\setminus ij}) = c_{ij} Q_1(e_{ij}) \prod_{r=1}^{q} \prod_{\substack{e_{ij} \subseteq \boldsymbol{e}_{S'} \\ \boldsymbol{e}_{S'} \in S_{k_r}^{ll'}}} Q_{S'}^{(k_r,2)}(\boldsymbol{e}_{S'}), \tag{65}$$

where $\{k_r\}_{1\leq r\leq q}$ is a subset of $\{2, .., p\}$. $S_{k_r}^{ll'}$ is the subset of S_k such that if $\boldsymbol{e}_S \in S_{k_r}^{ll'}$ then for all $e_{i_2 j_2} \in e_{\boldsymbol{S}}$, we have that $\{i_2, j_2\} \cap U_l \neq \emptyset$ and $\{i_2, j_2\} \cap U_{l'} \neq \emptyset$.

In general, it is not possible to include a network structure H_k in Formulas (65) whenever there exists a sequence $\{i_1, ..., i_{2l+1}\}$ such that $i_1 = i_{2l+1}$, $i_w \in V(H_k)$ (for all $w = 1, ..., 2l + 1$)and $e_{i_w i_{w+1}}$ belong to the relational variables determined by H_k. For instance, the network structure two path is possible to include but not transitive triads.

The set of Formulas (64) and (65) states that the conditional distribution of relational variables can be partitioned into different blocks. Each block reflects an agents' decision to update relations while only considering a subset of variables, and it represents the likelihood of agents to update variables in such a way that they are in hyponetworks with a specific structure.

For $k = 1$, structure H_1 represents the likelihood of an agent to be connected with another agent and this likelihood is influenced by the group membership of both agents.

For H_1,

$$Q_1(\boldsymbol{e}_{ij}) = \begin{cases} c_1^S & e_{ij} = 1 \text{ and } \{i,j\} \subset U_l \text{ for some } l \\ 1 - c_1^S & e_{ij} = 0 \text{ and } \{i,j\} \subset U_l \text{ for some } l \\ c_0^S & e_{ij} = 1 \text{ and } \{i,j\} \not\subset U_l \text{ for all } l \\ 1 - c_0^S & e_{ij} = 0 \text{ and } \{i,j\} \not\subset U_l \text{ for all } l \end{cases} \tag{66}$$

If $c_0^S < c_1^S$, then communities are assortative and H_1 accounts for the tendency of agents to prefer connections with agents in the same community. Similarly, when $c_0^S > c_1^S$ communities are disassortative and H_1 accounts for the tendency of agents to prefer connections with agents in other communities.

Assumption 7. *For $k > 1$ and $i \in \{1, 2\}$,*

$$Q_S^{(k,i)}(\boldsymbol{e}_S) = \begin{cases} c^S_{(k,i)} & \textit{if } \boldsymbol{e}_S \sim H_k \\ 1 - c^S_{(k,i)} & \textit{otherwise} \end{cases} \tag{67}$$

for some $c^S_{(k,i)} \in (0, 1) \setminus \{\frac{1}{2}\}$.

When $c^S_{(k,i)} > \frac{1}{2}$, this tendency states that agents are more likely to update e_{ij} in such a way that it adds a hyponetwork with structure H_k, and when $c^S_{(k,i)} < \frac{1}{2}$ agents are more likely to update e_{ij} in a manner that it does not add a hyponetwork with structure H_k. When updating e_{ij} cannot contribute to add a hyponetwork with structure H_k, this tendency does not make more or less likely to add a relation.

Assumption 6 asserts that the social mechanisms underlying the creation or sever of a relation between two agents depend only on whether both agents belong to the same community or not. Further, it states that the social mechanisms activated in the updating process of relations between communities are a subset of the social mechanisms activated in the relational process within communities (this point can be relaxed).

The three previous assumptions are set at the microscopic level as they describe the dynamics at the agent-agent interaction level. These assumptions plus another technical assumption define a particular class of ERGM with community structure.

Theorem 8. *If Assumptions 5, 6 and 7 are satisfied and any realisation has positive probability, then*

$$P_{\boldsymbol{\theta}^S}(\boldsymbol{e}_S) = \frac{1}{c_S(\boldsymbol{\theta}^S)} \exp\left(\sum_{k=1}^{p+q} \theta_k^S \Gamma_k(\boldsymbol{e}_S) \right). \tag{68}$$

When $k \in \{1, ..., p\}$, $\theta_k^S = \log\left(\frac{c^S_{(k,1)}}{1-c^S_{(k,1)}}\right)$ and $\Gamma_k(\boldsymbol{e}_S)$ is the number of hyponetworks in $\boldsymbol{e}_S$

isomorphic to H_k with $V(\boldsymbol{e}_S) \subseteq U_l$ for some l.

When $k \in \{p+1, ..., p+q\}$, we let $r \equiv k(\mod p)$, $\theta_k^S = \log\left(\frac{c^S_{(r,2)}}{1-c^S_{(r,2)}}\right)$ and $\Gamma_k(\boldsymbol{e}_S)$ is the number of hyponetworks in $\mathbf{e}_S$ isomorphic to H_{k_r} such that there exist two communities U_l and $U_{l'}$ satisfying that for all $e_{ij} \in e_{\boldsymbol{S}}$, we have that $\{i,j\} \cap U_l \neq \emptyset$ and $\{i,j\} \cap U_{l'} \neq \emptyset$.

$c_S(\boldsymbol{\theta}^S)$ is a normalising constant,

$$c_S(\boldsymbol{\theta}^S) = \sum_{\bar{\boldsymbol{e}}_S \in \Omega_S} \exp\left(\sum_{k=1}^{p+q} \theta_k^S \Gamma_k(\bar{\boldsymbol{e}}_S)\right).$$

Ω_S denotes the set of all the possible realisations of the variables $\mathbf{e}_S$. The dimension of $\boldsymbol{\theta}^{\boldsymbol{S}}$ is equal to $p+q$.

We will sometimes interchange the notation Γ_k with $\Gamma_{k,U}$ for $k = 1, .., p$ ($\Gamma_{k,U}$ denotes a network statistics defined by relations within communities) and with $\Gamma_{r,W}$ denotes the network statistics defined by the relations between communities). For $n < B$, it is possible that all agents belong to the same community, in which case $\Gamma_{r,W}$ are removed from Formula (68).

Notice that by knowing the joint distribution, we can determine when two variables are independent. For instance, when two agents i and j belong to the same community, variable e_{ij} is dependent to at most $B(B-1)$ variables, while when they belong to different communities e_{ij} is dependent to at most $2B^2$. The dependency structure between variables can be described by a graph, in which nodes are variables and two nodes are connected if the associated variables are dependent. Here, we call the pair $G(V, E)$ a graph when V represents an abstract set of elements (nodes) and the set of links between the nodes, E. In general, we define the dependency graph as follows (Féray et al., 2016).

Definition 9. *Given a set of random variables $(Y_v)_{v \in V}$; we say that a graph $D(V, E)$ is the dependency graph of the variables if the following holds. For any disjoint subsets V_1, V_2 of V such that there is no edge between the two subsets in D, the set of variables $(Y_v)_{v \in V_1}$ and $(Y_v)_{v \in V_2}$ are independent.*

The upper bound for the size of the communities in Assumption 5 is needed to control for the rate of convergence of the maximum degree of the dependency graph, which is a sufficient condition for the central limit theorem to hold (Féray et al., 2016).

Another representation of the dependency structure between the relational variables is possible with the help of the mathematical framework of Markov random fields (also known as graphical models) (Murphy, 2012). A Markov random field (MRF) consists of an undirected graph $D_2(V, E)$ where each node is associated with a random variable and an edge between a pair of nodes encodes the dependency between the corresponding variables. For a MRF, any variable $v \in V$ satisfies the local Markov property, which says that Y_v is conditionally independent of any other variable given all its neighbours:

$$Y_v \perp\!\!\!\perp Y_{V \setminus N \cup \{v\}} \mid Y_{N(v)}$$

where $N(v)$ is the set of neighbours of Y_v in D_2.

Markov random fields are the mathematical framework underlying exponential random graph, whenever ERGM are formulated with assumptions at the agent-agent interaction level. According to the Hammersley-Cliford theorem, the joint distribution of a MRF is a Gibbs distribution (Besag, 1974), and this theorem is used to prove Theorem 8. The next result shows the relationship between ERGM and MRF.

Result 9. *Let $\boldsymbol{e}_S$ be a set of relational variables satisfying Assumptions 6 and 7. If we define the graph $D_2(V, E)$ as follows: 1) $V = \boldsymbol{e}_S$, and 2) $(e_{ij}, e_{i_2 j_2}) \in E$ if and only if there exists a function $Q_{S'}^{(k,i)}$ such that $ij, i_2 j_2 \in S'$, then the graph $D_2(V, E)$ is a Markov random field.*

In order to be consistent with the notation introduced by Frank and Strauss (1986), we refer to the graph constructed in Result 9 as the dependence graph. The dependency graph in Definition 9 and the dependence graph constructed in Result 9 are used in two different context. The first type of graphs is used to show that the central limit theorem holds (Féray et al., 2015a) while the second type of graphs is used for determining the joint distribution of random variables (Frank and Strauss, 1986; Pattison and Robins, 2002).

6.3 Data assumptions

We define a sequence $\{\mathcal{F}_n\}_{n=2}^{\infty}$ of subsets of $\mathcal{F}$ such that if $i, j \in V(\boldsymbol{e}_{\mathcal{F}_n})$, then $ij \in \mathcal{F}_n$, and the cardinality of $V(\boldsymbol{e}_{\mathcal{F}_n})$ equals n. $\{\boldsymbol{e}_{\mathcal{F}_n}\}_{n=2}^{\infty}$ represents an increasing sequence of random networks ($\boldsymbol{e}_{\mathcal{F}_n} \subset \boldsymbol{e}_{\mathcal{F}_{n+1}}$) with a growing number of agents. For the sake of simplicity, we denote $\boldsymbol{e}_{\mathcal{F}_n}$ and $\boldsymbol{\theta}^{\mathcal{F}^n}$ by $\boldsymbol{e_n}$ and $\boldsymbol{\theta^n}$, respectively. If Ω_n denotes the set of all possible realisations of $\boldsymbol{e_n}$, then the sequence $\{(\Omega_n, 2^{\Omega_n}, P_{\boldsymbol{\theta^n}})\}_{n=2}^{\infty}$ defines a sequence of ERGM with probability distributions defined by the same network statistics, but it is possible that the parameters are functions of the number of agents in the network. For instance, we may assume that the social mechanisms underlying the formation of networks is a constant function of the number of agents, but their effects may vary.

Our next problem is to find a sequence of estimators $\{\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}\}_{n=2}^{\infty}$ to infer some aspects of $\{\boldsymbol{\theta^n}\}_{n=2}^{\infty}$; this can be achieved via point estimation, confidence intervals, significance test, hypothesis testing. In the simplest case when $\boldsymbol{\theta^n}$ is constant on n, a desirable large-sample property of estimators is consistency, i.e. converges to the true parameter.

6.3.1 Estimation and p-values

An estimator for the parameter $\boldsymbol{\theta^n}$ in Equation (68) is the global maximum of the $\log$ likelihood function $l(\boldsymbol{\theta^n}; \boldsymbol{e_n}) = \log\left(P_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{e_n})\right) = \sum_{k=1}^{p+q} \theta_k^n \Gamma_k(\boldsymbol{e_n}) - \log(c_n(\boldsymbol{\theta^n}))$. This estimator is denoted by $\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}$ and it is known as the maximum likelihood estimator (MLE).

$$\hat{\boldsymbol{\theta}}^{\boldsymbol{n}} = \arg\max_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^{p+q}} l(\tilde{\boldsymbol{\theta}}; \boldsymbol{e_n}) \tag{69}$$

For distributions of the discrete exponential family, like the one presented in Formula (68), MLE satisfies several important properties (Barndorff-Nielsen, 2014). First, when a global maximum exists and is unique, the expected value of the network statistics evaluated at the

For the undirected case, we assume that if $i, j \in V(\boldsymbol{e}_{\mathcal{F}_n})$, then either $ij \in \mathcal{F}_n$ or $ji \in \mathcal{F}_n$.

In general, the non-constant function of the parameters on the number of agents in the network often complicates the notion of consistency of estimators (Strauss and Ikeda, 1990). Consistency for non-constant parameters is treated in (Chandrasekhar and Jackson, 2014).

MLE equals the observed statistics, i.e. $\mathrm{E}_{\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}}(\boldsymbol{\Gamma}) = \boldsymbol{\Gamma}(\boldsymbol{e_n})$. Second, the Hessian of the log-likelihood function evaluated at the MLE equals minus the Fisher Information Matrix at the MLE, $\nabla^2 l(\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}; \boldsymbol{e_n}) = -I(\hat{\boldsymbol{\theta}})$. Further, the Fisher information matrix satisfies $I(\hat{\boldsymbol{\theta}}) = \mathrm{Cov}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\Gamma})$.

The case of non-existence of a global maximum for distributions in the discrete exponential family is a well-known problem (Barndorff-Nielsen, 2014). It is known that if we define the set

$$\mathcal{C}_n = \{x \in \mathbb{R}^{p+q} : x = \boldsymbol{\Gamma}(\bar{\boldsymbol{e}}_{\boldsymbol{n}}) \text{ with } \bar{\boldsymbol{e}}_{\boldsymbol{n}} \in \Omega_n\}$$

and $Conv(\mathcal{C}_n)$ denotes the convex hull of $\mathcal{C}_n$; then a global maximum for $\bar{\boldsymbol{e}}_{\boldsymbol{n}}$ exists if and only if $\boldsymbol{\Gamma}(\bar{\boldsymbol{e}}_{\boldsymbol{n}}) \in Conv^{\circ}(\mathcal{C}_n)$, i.e. the observed statistics are in the interior of the convex hull of $\mathcal{C}_n$. Intuitively, when the true parameters are a constant function of n and the MLE is consistent, then we would expect that the non-existence issue of the global maximum vanishes as n tends to infinity. Following the rationale behind the MLE for regression models (Hayashi, 2000) whenever the global maximum does not exist, we assign an arbitrary value to $\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}$ so that the MLE is well-defined for any sample.

For the exponential random graph model, a key element for constructing p-values with network data is the fact that $I^{\frac{1}{2}}(\bar{\boldsymbol{\theta}}^{\boldsymbol{n}})(\hat{\boldsymbol{\theta}}^{\boldsymbol{n}} - \boldsymbol{\theta}^{\boldsymbol{n}})$ converges to a normal distribution if and only if $\mathrm{Cov}_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}^{-\frac{1}{2}}(\boldsymbol{\Gamma})(\boldsymbol{\Gamma}^{\boldsymbol{n}} - \mathrm{E}_{\boldsymbol{\theta}^{\boldsymbol{n}}}(\boldsymbol{\Gamma}))$ converge to a normal distribution for some $\bar{\boldsymbol{\theta}}^{\boldsymbol{n}} = \lambda\hat{\boldsymbol{\theta}}^{\boldsymbol{n}} + (1-\lambda)\boldsymbol{\theta}^{\boldsymbol{n}}$ with $\lambda \in (0,1)$. When the parameters are constant on n, it is possible to substitute $\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}$ by $\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}$, as we show later in Result 20. However, up to our knowledge, it is unclear why the substitution is still valid when the absolute value of the parameters tends to infinity.

The class of ERGM satisfying Assumptions 5-8 are called community exponential random graph model (C-ERGM).

Assumption 8.

$$\lim_{n\to\infty} \inf \sigma_n^2(k) = \frac{\mathrm{Var}_{\boldsymbol{\theta}^n}(\Gamma_k)}{n_k} \to \sigma^2 > 0$$

If we define the score function $s(\boldsymbol{\theta}^{\boldsymbol{n}}; \boldsymbol{e_n}) = \nabla l(\boldsymbol{\theta}^{\boldsymbol{n}}; \boldsymbol{e_n})$, then the Fisher information matrix is defined as $\mathrm{Cov}_{\boldsymbol{\theta}^{\boldsymbol{n}}}(s(\boldsymbol{\theta}^{\boldsymbol{n}}; \boldsymbol{e_n}))$.

Our main theorem gives conditions under which test statistics are approximately normal distributed for the class C-ERGM.

Theorem 9. *If the Assumptions of Theorem 8 and Assumption 8 are satisfied, then for all* $k \in \{1, ..., p+q\}$

$$\frac{\Gamma_k(\boldsymbol{e_n}) - \mathrm{E}_{\boldsymbol{\theta}^{\boldsymbol{n}}}(\Gamma_k)}{\sqrt{\mathrm{Var}_{\boldsymbol{\theta}^{\boldsymbol{n}}}(\Gamma_k)}} \xrightarrow{D} \mathcal{N}(0,1) \tag{70}$$

Assuming infinite communities is possible, but we require an assumption of the size of the largest observed community.

Assumption 9. *For* $\boldsymbol{e_n}$ *and its associated agents* $V(\boldsymbol{e_n})$*, let* $U_{l^\star,n}$ *denote the largest observed community, i.e.* $|V(\boldsymbol{e_n}) \cap U_{l^\star,n}| \geq |V(\boldsymbol{e_n}) \cap U_l|$ *for all communties. Let us assume that the cardinality of* $U_{l^\star,n}$ *is* $o((n)_k)$ *for all network statistics.*

This assumption is not related to the properties of the agents but on the data-generating process. For determining the validity of this assumption, a question to be addressed is if the communities are assumed to be infinite, how should network data be collected, so that the largest observed community satisfies Assumption 9.

As it was mentioned before, for a large class of social networks, it is expected that the parameters are a non-constant function of the cardinality of $V(\boldsymbol{e_n})$. Proving the general case is possible using Theorem 14 in (Féray et al., 2015a). In fact, an upper bound for the convergence rate to the normal distribution with respect to the Kolmogorov distance is obtained. Where for two random variables X, Y, the Kolmogorov distance (d_K) between them is defined as $d_K(X,Y) = \sup_{a\in\mathbb{R}} \Big|P(X \leq a) - P(Y \leq a)\Big|$.

Theorem 10. *If the Assumptions of Theorem 9 hold, then*

$$d_K\Big(\frac{\Gamma_k(\boldsymbol{e_n}) - \mathrm{E}_{\boldsymbol{\theta}^{\boldsymbol{n}}}(\Gamma_k)}{\sqrt{\mathrm{Var}_{\boldsymbol{\theta}^{\boldsymbol{n}}}(\Gamma_k)}}, \mathcal{N}(0,1)\Big) \leq \frac{56}{\sigma}\sqrt{\frac{D_{n,k}}{n_k}} \tag{71}$$

where $D_{n,k}$, n_k denote the maximum degree and the number of nodes of the dependency graph $\boldsymbol{G}_{n,k}$, respectively.

If Assumption 9 holds, then $D_{n,k} = o((n)_k)$ and the left side of Equation (71) converges to zero.

6.4 Network model assumptions

Networks are often modelled as the joint distribution of relational variables, but they are usually formulated by making assumptions at the agent-agent interaction level (microscopic-level). Assumptions at the microscopic level, like the one formulated in Assumption 6, are intended to describe how the updating process of each relational variable is influenced by existence and non-existence of relations between the agents. Thus, the work by Besag (1974) to determine the joint distribution given conditional distributional assumptions has proven to be crucial for the development of network models.

In this section, we present the underlying assumptions of some subclasses of ERGM. We start from the most simple model (stochastic block models) and we finished with a formulation of ERGM in terms of rational agents.

6.4.1 Stochastic blockmodel

One of the simplest random graph models used to explain the formation and sever of relations between agents with a community structure are the stochastic block models. In a stochastic blockmodel, the relational variables are independent and the probabilities to have a link between two agents is a function of the group to which the pair of agents belong (Holland et al., 1983).

In the simplest stochastic blockmodel (SBM), each agent is assigned to one of $K(< \infty)$ communities and the probability that a specific agent i relates to another agent j is given by

Formulations of ERGM which do not make assumptions at the agent-agent interaction level were presented by Wasserman and Pattison (1996) and Park and Newman (2004).

$P(e_{ij} = 1) = c_{U_l U_{l'}}$ when i is in community U_l and j is in community $U_{l'}$. This condition says that for any two agents in the same community, the choices they make are governed by the same probability distribution (Anderson et al., 1992). For the simplest SBM, the likelihood of an agent i to relate with agent j can be written as

$$Q_1(e_{ij}) = \begin{cases} c_{1,1} & \text{if } e_{ij} = 1,\, i \in U_l \text{ and } j \in U_l \\ 1 - c_{1,1} & \text{if } e_{ij} = 0,\, i \in U_l \text{ and } j \in U_l \\ c_{1,2} & \text{if } e_{ij} = 1,\, i \in U_l \text{ and } j \in U_{l'} \\ 1 - c_{1,2} & \text{if } e_{ij} = 0,\, i \in U_l \text{ and } j \in U_{l'} \end{cases}$$

For SBM, constructing valid p-values is feasible by removing the assumptions on the existence of a partition of the agents in finite communities (Assumption 5) but the independence between variables is strengthened. Let us assume that the variable e_{ij} is independent of all other variables, except possible from e_{ji}. The possible dependence between the two variables is modelled by the function Q_2 which controls the likelihood of agents to reciprocate.

Assumption 10.

$$P(e_{ij} | e_{S \setminus \{ij\}}) = Q_1(e_{ij}) Q_2(e_{ij}, e_{ji})$$

where

$$Q_2(e_{ij}, e_{ji}) = \begin{cases} c_2 & \textit{if } e_{ij} = e_{ji} = 1 \\ 1 - c_2 & \textit{otherwise} \end{cases}$$

Result 10. *Let Assumptions 8 and 10 be satisfied. If the distribution assigns positive probability to any realisation, then the joint distribution is*

$$P_{\boldsymbol{\theta}}(\boldsymbol{e_n}) = \frac{1}{c_n(\boldsymbol{\theta})} \exp\left(\sum_{k=1}^{2} \theta_{1,k} \Gamma_{1,k}(\boldsymbol{e_n}) + \theta_2 \Gamma_2(\boldsymbol{e_n}) \right). \tag{72}$$

where $\theta_{1,k} = \log(\frac{c_{1,k}}{1-c_{1,k}})$ *and* $\theta_2 = \log(\frac{c_2}{1-c_2})$.

If in addition the number of observed agents in each community tends to infinity, then the asymptotic normality in Formula (70) is satisfied.

Although for stochastic blockmodels, the construction of p-values does not need to assume a partition of the agents in finite communities, these models do not generate networks with a structure similar to the ones observed in reality (Karrer and Newman, 2011). This may be a result of the independence assumption between variables in this class of models, which does not seem plausible in network data (Ellwardt et al., 2012; Lorenz and Hall, 2013; Cranmer et al., 2012).

6.4.2 Markov random graph

From now on, unless otherwise stated, we will focus on undirected networks. Markov random graph is a class of network model proposed by Frank and Strauss (1986). These models assume that the existence of a relation from agent i to j does not provide any information on the updating process of the variable e_{kl} if it is known the value of all other variables and agents k, l are two different agents from i, j (Figure 14).

Assumption 11. *Two variables e_{ij} and e_{kl} are conditionally independent given all other variables if and only if $\{i, j\} \cap \{k, l\} = \emptyset$.*

As an example, when the set of variables illustrated in Figure 14 **(a)** satisfy Assumption 11, the dependence graph equals the graph shown in Figure 14 **(b)**

The assumption of Markov graph models defines a class of probability distributions with 9 different network statistics represented by hyponetworks with less than four agents and less than four relational variables (Figure 15). The possible number of network statistics grows linearly as the number of agents in the network increases. As a result, it is custom to set several parameters to zero. When all parameters are set to zero except for the parameter of the network statistic number of relations, the result is the Bernoulli graph models. Markov graph models are often restricted to include the network statistics: the number of relations, number of two-path and number of triangles, see Figure 13 **(a)**-**(c)**.

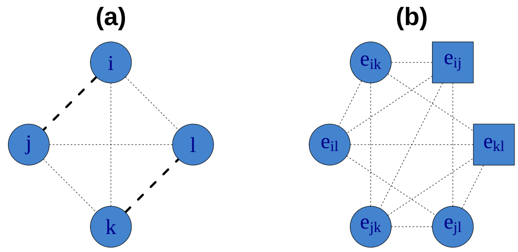


Figure 14: In **(a)**, if Assumption 11 holds, then the relational variables e_{ij} and e_{kl} are conditionally independent given $(e_{ik}, e_{jk}, e_{il}, e_{jl})$. The dependence graph of the model is illustrated in **(b)**, with the nodes e_{ij} being connected to all other nodes except with e_{kl}.

6.4.3 Partial conditional dependence assumption

A larger class of ERGM was proposed by Pattison and Robins (2002) with their partial conditional dependence assumption. Here, two variables might be independent conditional on the rest of the variables if certain relations in the network do exists. Snijders et al. (2006) proposed that two variables e_{ij}, e_{kl} are conditional dependent given the rest of the variables if they are adjacent (as in Markov graph models) or if $e_{ik} = e_{kl} = 1$ (if the variables could potentially be part of a four cycle). These assumptions allow to incorporate to models the network statistics number of k-triangles (see Figure 13 **(e)**), which is not possible with Markov graph models.

A k-triangles is defined by the existence of a relation between two agents i, j and the existence of k other agents related to both i and j. k-triangles express transitive-alike concept with the set of agents of cardinality larger than 3 (Snijders et al., 2006). The network statistics number of k-triangles T_k in a network equals

$$T_k = \sum_{l=1}^{n-2} \binom{l}{k} P_l$$

where P_l equals the number of relations between any pair of agents i, j with $e_{ij} = 1$ such that both agents are related to exactly l other agents, see Figure 13 (**e**).

The intuition of partial conditional dependence is formalised with the concept of conditional independence.

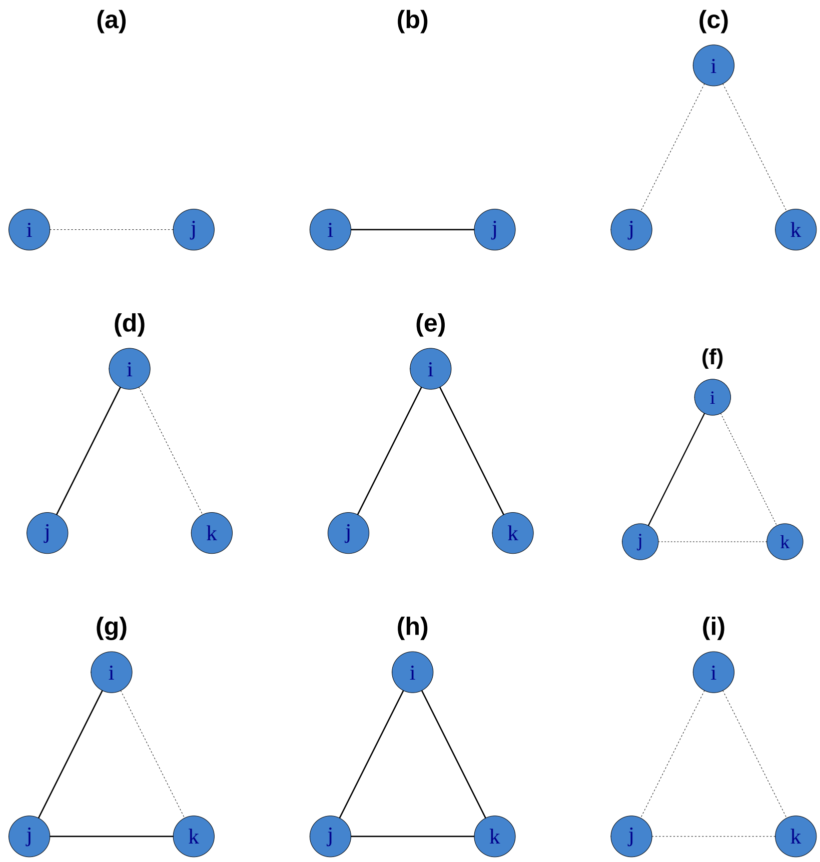


Figure 15: Hyponetworks allowed in Markov random graph models with less than four agents and less than four relational variables. A solid line between two agents indicates a relation between them while a dot the line indicates that there is no relation between them. The absence of a line represents that there is no variable defined between the two agents.

Definition 10. *We say that two relational variables e_{ij} and e_{kl} are conditionally independent on a realisation $\bar{e}_B$ of the variables e_B with $e_{ij}, e_{kl} \notin e_B$ if*

$$P(e_{ij}, e_{kl}|e_B = \bar{e}_B) = P(e_{ij}|e_B = \bar{e}_B)P(e_{kl}|e_B = \bar{e}_B) \tag{73}$$

In other words, two variables e_{ij}, e_{kl} are conditional independent on the realisation $\boldsymbol{e}_B = \bar{\boldsymbol{e}}_B$, if given the knowledge that $\boldsymbol{e}_B$ equals $\bar{\boldsymbol{e}}_B$, then knowing the value of the variable e_{kl} provides no information on the probability of observing a relationship between k and l.

We say that the set of relational variables $\boldsymbol{e}$ satisfies the *partial conditional dependence* *assumptions* if the following two assumptions hold.

Assumption 12. *The dependence graph, D, of $\boldsymbol{e}_S$ is a Markov random field and the graph D is known.*

Assumption 13. *For any pair of variables and any realisation $\bar{e}_B$, we know if they are conditionally independent on the realisation $\bar{e}_B$.*

The Markov random graph models satisfy Assumption 12, and Assumption 13 holds when two variables e_{ij} and e_{kl} are conditionally independent if $e_{ik} = e_{jk} = e_{il} = e_{jl} = 0$. In particular, the proposed assumptions to justify k-triangles implies that the dependence graph is the complete graph, and the Hammersley-Clifford theorem alone is insufficient to describe the distribution over the networks. Following the work by Pattison and Robins (2002), the next two results show some relations of the parameters and the partial conditional dependence assumption.

Before presenting the next two results, we need a few new concepts. Let us assume that the set of variables $\boldsymbol{e}_S$ satisfy assumption 12 and 13 and let D be the associated dependence graph. We denote by $\mathcal{C}$ the set of all cliques in D, where the set of nodes $\boldsymbol{e}_{S'}$ in D form a clique if any pair $e_{ij}, e_{kl} \in \boldsymbol{e}_{S'}$ are connected. In Figure 14 **(b)**, $\{e_{ik}\}$, $\{e_{ik}, e_{ij}\}$ and $\{e_{ik}, e_{ij}, e_{il}\}$ are cliques, and there is no cliques with more than three nodes. For any set of variables $\boldsymbol{e}_B$, we define a relation, $\overset{B}{\sim}$, over the cliques as follows: $\boldsymbol{e}_{S'} \overset{B}{\sim} \boldsymbol{e}_{S''}$ if and only if $\boldsymbol{e}_{S'} \triangle \boldsymbol{e}_{S''} \subset \boldsymbol{e}_B$. $[\mathcal{C}]_B$ denotes the partition induced by $\overset{B}{\sim}$ and $[\boldsymbol{e}_{S'}]$ is the class with representative element $\boldsymbol{e}_{S'}$. Back to Figure 14 **(b)**, when $\boldsymbol{e}_B = \{e_{il}, e_{ij}\}$, then $\{e_{il}, e_{ij}, e_{ik}\} \overset{B}{\sim} \{e_{ik}\}$. In general, if $e_{ij} \in \boldsymbol{e}_{S'} \setminus \boldsymbol{e}_B$ and $\boldsymbol{e}_{S'} \overset{B}{\sim} \boldsymbol{e}_{S''}$, then $e_{ij} \in \boldsymbol{e}_{S''}$.

Result 11. *If Assumptions 12 and 13 are satisfied, then the probability distribution is of the form*

$$P(\boldsymbol{e}) = \frac{1}{c(\boldsymbol{\theta})} \exp\Big(\sum_{\boldsymbol{e}_{S'} \in \mathcal{C}} \theta_{S'} \prod_{e_{i'j'} \in \boldsymbol{e}_{S'}} e_{i'j'}\Big) \tag{74}$$

such that

$$\sum_{\boldsymbol{e}_{S'} \in [\boldsymbol{e}_C]_{ij,kl}} \theta_{S'} \Big(\prod_{e_{i'j'} \in \boldsymbol{e}_{S'} \cap e_B} \bar{e}_{i'j'} \Big) = 0 \tag{75}$$

for all $[\boldsymbol{e}_C]_{ij,kl}$ *such that there exits* e_{ij}*,* $e_{kl} \in [\boldsymbol{e}_C]_{ij,kl}$ *and* e_{ij}*,* e_{kl} *are conditional independent on* $\bar{\boldsymbol{e}}_B$*.*

A particular case of the last result occurs when the realisations $\bar{\boldsymbol{e}}_B$ are vectors of zeros, $\bar{\mathbf{0}}$.

Result 12. *If Assumptions 12 and 13 are satisfied and the realisations are of the form* $\bar{\boldsymbol{e}}_B = \bar{\mathbf{0}}$*, then Formula (75) holds with the additional constraints that* $\theta_{S'} = 0$ *if (i) there exists* e_{ij}*,*e_{kl} *conditionally independent on* $\bar{\mathbf{e}}_B$ *and (ii)* $e_{ij}, e_{kl} \in \boldsymbol{e}_{S'}$ *with* $\boldsymbol{e}_{S'} \cap \boldsymbol{e}_B = \emptyset$*.*

Although, Results 11 and 12 narrow the number of parameters different from zero; the number of network statistics allowed in the model is too large for including all. Thus, as in Markov random graph models, it is custom to set to zero most of the parameters. Alternatively, since the value of the parameters are highly dependent, new network statistics have been defined by imposing constraints on the parameter space. For instance, the network statistics geometrically weighted edge-wise shared partner (GWESP) was defined by constraining the value of the parameters for the network statistics number of k-triangles (T_k) (Snijders et al., 2006; Hunter and Handcock, 2006).

$$u(\boldsymbol{e}, \tau) = 3T_1(\boldsymbol{e}) - \frac{T_2(\boldsymbol{e})}{\exp{(\tau)}} + ... + (-1)^{n-3} \frac{T_{n-2}(\boldsymbol{e})}{\exp{((n-3)\tau)}}$$

Similarly, the geometrically weighted degree counts was proposed by constraining the parameters associated with the statistics k-stars (Snijders et al., 2006). Since the introduction of GWESP, GWESP has substituted the use of triangles in empirical studies (Ilany et al., 2013; McFarland et al., 2014; Song, 2015).

6.4.4 Potential games and exponential random graph model

The parallels observed between potential games and Markov random field (Babichenko and Tamuz, 2016) illustrates another formulation of ERGM (Mele, 2013). Here, the network formation is the result of each agent maximising its utility function. As before we assume a set of agents V with their respective relational variables $\boldsymbol{e}_S$. $E_i = \{(\bar{e}_{ij})_{j \neq i} : \bar{e}_{ij} \in \{0, 1\}\}$ is the set of all strategies that are available to agent i and each agent has an utility function $u_i : \mathcal{G}^n \to \mathbb{R}$. Since, we assume that agents decide their outgoing relations, in this subsection we consider directed network.

Assumption 14. *Let us assume that each agent i has an utility function u_i satisfying*

$$u_i(\boldsymbol{e}_S) = \sum_{k=1}^{p} \sum_{\boldsymbol{e}_S \in S_k} \theta_k \mathbb{1}_{\left\{ \substack{i \in V(\boldsymbol{e}_S) \\ \boldsymbol{e}_S \sim H_k} \right\}} \tag{76}$$

The utility of an agent is a linear function of its preferences to belong to hyponetworks with particular structures. When $\theta_k > 0$, agents prefer to belong to hyponetworks with structure H_k while when $\theta_k < 0$ agents prefer not to belong to hyponetworks with structure H_k. The network games defined by $(\boldsymbol{e}_1, ..., \boldsymbol{e}_{|V|}, u_1, ..., u_{|V|})$ with the utility functions satisfying Assumption 14 belong to the class of potential games. In Potential games, agents' incentives to change their strategy is expressed by a potential function Φ, and the Nash Equilibria correspond to local maxima of the potential function, $\Phi(\boldsymbol{e}_S)$.

Definition 11. *A game is an exact potential game if there exists a function $\Phi : \mathcal{G}^n \to \mathbb{R}$ with*

$$\Phi(\bar{e}_i, \bar{e}_{-i}) - \Phi(\tilde{e}_i, \bar{e}_{-i}) = u_i(\bar{e}_i, \bar{e}_{-i}) - u_i(\tilde{e}_i, \bar{e}_{-i})$$

for any agent i an for any realisations $\bar{e}_i$, $\bar{e}_{-i}$, $\tilde{e}_i$.

Definition 12. *A game is an ordinal potential game if there exists a function $\Phi : \mathcal{G}^n \to \mathbb{R}$ with*

$$\Phi(\bar{e}_i, \bar{e}_{-i}) - \Phi(\tilde{e}_i, \tilde{e}_{-i}) > 0 \textit{ if and only if } u_i(\bar{e}_i, \bar{e}_{-i}) - u_i(\tilde{e}_i, \tilde{e}_{-i}) > 0$$

for any agent i an for any realisations $\bar{e}_i$, $\bar{e}_{-i}$, $\tilde{e}_i$.

Result 13. *If Assumption 14 is satisfied, the game is an exact potential game with potential function equal to*

$$\Phi(\boldsymbol{e}_S) = \sum_{k=1}^{p} \theta_k \Gamma_k(\boldsymbol{e}_S)$$

There are two common dynamics to the updating process of strategies in the literature. In both an agent i is chosen at random, and 1) a relational variable e_{ij} is chosen to be updated or 2) agent i updates the relational variable $\boldsymbol{e}_i$.

Assumption 15. *Let us assume a discrete time and either one of the two things happen:*

1. (a) *At each time t a variable e_{ij} is chosen with probability p_{ij} and agent i updates the relation.*

   (b) *For all ij, $p_{ij} = \frac{2}{|V|(|V|-1)}$.*

2. (a) *At each time t a set of variables e_i is chosen with probability p_i and agent i updates the relation.*

   (b) *For all i, $p_i = \frac{1}{|V|}$*

Assumption 16. *The updating process is driven by maximising their utility.*

Under Assumptions 15.2 and 16, the game converge to a Nash Equilibria in a finite number of steps. The convergence to a Nash Equilibria does not necessarily occur under Assumption assumptions 15.1. For instance, when $\boldsymbol{e}_S = (e_{ij}, e_{ji}, e_{jk}, e_{kj}, e_{ik}, e_{ki})$, and the gains to create relations is -1 and the gain to be related to the other two agents is 2. Then, when the dynamics start from the null networks, and relations are updated according to Assumption 15.1, agents always choose not to create a relation, since the agents utility decreases from 0 to -1. However, the only Nash Equilibrium of the game is when all agents choose to relate to the other agents.

In case 15.1, we can think that an agent updates only one relation at a time, while in case 15.2 we can imagine that each time an agent is chosen and it has the opportunity to update all its relations.

The selection of the best strategy at each period is a strong assumption on the agents' rationality. A relaxation of this assumption is to assume that agents have bounded rationality, and if an agent were going to update a relation, it would take suboptimal decisions. We do it by transforming u_i by monotonic functions that preserves the order of agents' preference and existence of a potential function.

Result 14. *Let v_i^β be defined by $v_i^\beta(\boldsymbol{e}_S) = \frac{1}{c(\boldsymbol{\theta})}\exp(\beta u_i(\boldsymbol{e}_S))$ with $\beta > 0$. The game defined by the utility functions v_is is an ordinary potential game with potential function.*

$$P_{\boldsymbol{\theta}^\beta}(\boldsymbol{e}_S) = \frac{1}{c(\boldsymbol{\theta}^\beta)}\exp\big(\beta\sum_{k=1}^{p}\theta_k\Gamma_k(\boldsymbol{e}_S)\big)$$

Notice that a local maxima of $\Phi(\boldsymbol{e}_S)$ is a local maxima of $P_{\boldsymbol{\theta}^\beta}(\boldsymbol{e}_S)$ for any $\beta \in (0,\infty)$.

The next assumption impose bounded rationality to the agents.

Assumption 17. *If an agent i had to update the variable e_{ij} given the rest of the relations, then the agent would do it as follows*

$$P(e_{ij} = \bar{e}_{ij}|\boldsymbol{e}_{-ij} = \bar{e}_{-ij}) = \frac{v_i^\beta(\bar{e}_{ij}, \bar{e}_{-ij})}{v_i^\beta(\tilde{e}_{ij}, \bar{e}_{-ij}) + v_i^\beta(\bar{e}_{ij}, \bar{e}_{-ij})} \tag{77}$$

with $\tilde{e}_{ij} = 1 - \bar{e}_{ij}$.

Assumption 17, like Assumption 6, describes the updating process of relations at the agent-agent interaction level. The parameter β models how likely are agents to select suboptimal actions. When $\beta \to 0$, agents select any action with equal probability; and when, $\beta \to \infty$, agents select their best response. The following result shows that when agents make suboptimal decisions, then the limiting distribution over $\boldsymbol{e}_S$ is an ERGM.

Result 15. *If Assumptions 14, 15 and 17 are satisfied, then the limiting distribution of the stochastic process is*

$$P_{\boldsymbol{\theta}^\beta}(\boldsymbol{e}_S) = \frac{1}{c(\boldsymbol{\theta}^\beta)}\exp\big(\beta\sum_{k=1}^{p}\theta_k\Gamma_k(\boldsymbol{e}_S)\big)$$

When $\beta \to 0$ the probability over the networks converge to a uniform distribution, but by the principle of maximum multiplicity and when n is large, the probability over the number of relations is concentrated on a few values. When $\beta \to \infty$, all the weight of the probability distribution goes on the realisations of $\boldsymbol{e}_S$ that maximise the potential function $\Phi(\boldsymbol{e}_S)$. For large β, unimodal distributions for $\Phi(\boldsymbol{e}_S)$ occur in the degenerate case, when there is only one class of Nash equilibrium.

6.5 Dynamic process and data assumptions

Theorem 15 shows that it is possible to justify an ERGM as the limiting distribution of the process of updating relationships between a set of agents. This justification of ERGM was first presented by Snijders (2001) and the following result summarise previous findings.

Result 16. *Let $\mathbf{e}_n$ be a set of relational variables and let us assume that at each time t a set of variables is chosen according to Assumption 15. If at each period the updating of the chosen variables is consistent with Assumptions 5, 6 and 7 and π is a probability distribution over $\boldsymbol{e}_n$ at time zero, then the updating process defines a Markov chain $\{\mathbf{e}_n^\tau\}_{\tau\geq 0}$ such that*

$$\lim_{\tau\to\infty} \boldsymbol{Q}^\tau \pi = P_{\boldsymbol{\theta}} \tag{78}$$

where $\boldsymbol{Q}$ is the transition matrix of the Markov process and $P_{\boldsymbol{\theta}}$ is a probability distribution like the one presented in Equation (68).

In friendship networks between students in classrooms, it is plausible that only a few students have met before the first day of class. In this case, the initial distribution, π, puts higher probability on networks with a few number of relations. When analysing network data with ERGM, the initial distribution is exogenous and it is not assumed to be known, but it is (im-

We focus on the simplified version of discrete time. An example of the use of discrete time is presented in the temporal exponential random graph model (Hanneke et al., 2010).

plicitly) assumed that the time when observations are collected is sufficiently large, i.e. the probability of observing a network is not influenced by the initial condition. Alternatively, we could justify ERGM with stationarity by the following theorem.

Result 17. *Let the Assumptions of Result 16 be satisfied. If there exists $\tau \geq 0$ such that $\boldsymbol{e}_n^{\tau}$ has distribution $P_{\boldsymbol{\theta}}$, then for all $\tau_1 \geq 0$ and $\tau_2 \geq \tau$*

$$(\boldsymbol{e}_n^{\tau}, ..., \boldsymbol{e}_n^{\tau+\tau_1}) \stackrel{d}{=} (\boldsymbol{e}_n^{\tau_2}, ..., \boldsymbol{e}_n^{\tau_2+\tau_1})$$

A consequence of the theorem is that for any $\tau_2 \geq \tau$, $\boldsymbol{e}_n^{\tau_2}$ has distribution $P_{\boldsymbol{\theta}}$. The two previous results suggest that a difference between regression models and ERGM is that the analysis of network with ERGM makes assumptions on the time the data is collected.

6.6 Problems with statistical inference

In this section, we focus on estimation and inferential problems occurring in ERGM. First, we give a quick overview on the well-documented non-existence problem of MLE for ERGM and the convergence problem of Markov chain Monte Carlo methods for approximating MLE. Second, we discuss the violation of the central limit theorem (CLT) in ERGM, and we discuss the existing approaches to justify the CLT. Finally, we show that properties occurring at the community level do not necessarily occur at the macroscopic level.

Contrary to the established idea that estimation problems suggest misspecified models, we present an alternative interpretation, in which estimation problems suggest a small-sample problem, e.g. analysing a network with a single community. Since the remedies to solve misspecified problem may cause more harm when the true problem is a small-sample size problem, researchers need to consider both possibilities when interpreting a model.

6.6.1 Estimation problems

Apart from the non-existence of a global maximum of the log likelihood function, a challenge for ERGM and early-noted by Frank and Strauss (1986) is the intractability of computing exact ML estimators. For this reason, they proposed a maximum pseudolikelihood (MPL) which was extended for directed graph by Strauss and Ikeda (1990). However, the properties for MPL remain unknown; and MLE and MPL are seldom used in practice. Instead, Markov chain Monte Carlo maximum likelihood (MCMC-MLE) methods for approximating MLE are the common tool for estimating parameters (Byshkin et al., 2016). The first constructions of Monte Carlo algorithms to approximate MLE in ERGM followed the work by Geman and Geman (1984) and Geyer (1991).

Two cases of Markov graph models for undirected networks have proven to be valuable for understanding properties of ERGM and the convergence speed of MCMC-MLE: (1) the two-star model (defined by the network statistics number of relations and number of two-stars), and (2) the triangle model (number of relations and number of triangles). The two-star model, as well as the triangle model, are in many aspects similar to the well-known Curie-Weiss Ising models from statistical physics (Häggström and Jonasson, 1999; Park and Newman, 2004). The importance of these two classes of network models stems from the fact that they are simple models exhibiting phase transitions and bimodal distributions for the network statistics (Häggström and Jonasson, 1999; Park and Newman, 2004). When bimodal distributions occur in ERGM, it was shown that the dynamics of Monte Carlo algorithms to approximate MLE might take a long time to converge (Bhamidi et al., 2011; Snijders, 2002) and the MCMC-MLE may fail to converge even when the MLE exists (Handcock et al., 2003). As a result, estimated parameters using MCMC-MLE will not exist either because MLE does not exist or a poor mixing of the MCMC algorithm (Handcock et al., 2003).

A plausible reason for the non existence of the MLE (or "very large" MLE) in ERGM, and not discussed in the ERGM literature, is a small-sample-size problem, i.e. observing networks with a few communities. For the two star models defined by 4 agents, the argument that maximise the log-likelihood function exists only for three out of ten configurations of network statistics, see Figure 16. In logistic regression models, the problem of observing infinite estimates is known as the separation problem, and it is prone to occur in small samples (Albert and Anderson, 1984; Heinze and Schemper, 2002). The problem of infinite estimates in ERGM and logistic models is a broader problem and it is inherited from the discrete exponential distributions (Jacobsen, 1989).

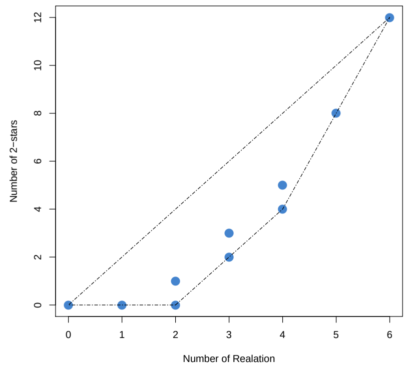


Figure 16: The dots represent possible values of the network statistics for a two-star model with four agent, and the dotted line indicates the convex hull of the points.

6.6.2 Counter-example of central limit theorem

Beyond Bernoulli graph models that assume independence between all variables, for the majority of ERGM used in empirical studies, any two relational variables are dependent. In these cases, when the relational variables are strongly dependent, a second problem stems when the assumptions rendering the central limit theorem for the network statistics are violated, and p-values are unreliable.

The next result shows that asymptotic normality of the network statistics does not hold for all ERGM. Similar to previous investigations of two-star and triangle models, the proof follows the approach of constructing an ERGM with properties that can be reduced to the Ising Weiss-Curie Potts models (see Supplementary Materials).

Result 18. *Let us define a sequence of ERGM by the network statistics number of relations (Γ_1) and number of two-stars (Γ_2) and their parameters are a function of the number of agents,* $c_1 = \frac{\exp(\theta_1)}{1+\exp(\theta_1)}$ *and* $c_2^n = \frac{\exp(\frac{\theta_2}{2n})}{1+\exp(\frac{\theta_2}{2n})}$.

Case 1. If 1) $\theta_2 = 4$ *and* $\theta_1 = -\theta_2$ *or 2)* $\theta_2 \geq 0$ *and* $\theta_1 > -\theta_2$*, then*

$$\frac{\Gamma_1(\boldsymbol{e_n}) - \mathrm{E}_{\boldsymbol{\theta}^n}(\Gamma_1)}{m^{\frac{3}{4}}} \rightarrow \exp\left(\frac{s^4}{12}\right)$$

where $m = \frac{n(n-1)}{2}$*, the number of relational variables.*

Case 2. If $\theta_2 < 4$ *and* $\theta_1 = -\theta_2$*, then*

$$\frac{\Gamma_1(\boldsymbol{e_n}) - \mathrm{E}_{\boldsymbol{\theta^n}}(\Gamma_1)}{\sqrt{\frac{m}{4-\beta}}} \rightarrow \mathcal{N}(0,1)$$

Case 3. If $\theta_2 > 4$ *and* $\theta_1 = -\theta_2$*, then*

$$\frac{\Gamma_1^n(\boldsymbol{e_n}) - \mathrm{E}_{\boldsymbol{\theta^n}}(\Gamma_1)}{m} \rightarrow \frac{1}{2}\delta_{u_1(\beta)} + \frac{1}{2}\delta_{u_2(\beta)}$$

where $u_1(\beta), u_2(\beta) \in (0,1)$.

The parameter θ_1 describes the idiosyncratic behaviour with $\theta_1 > 0$ expressing the tendency of agents to prefer closing a relationship when facing the decision to relate or not relate with another agent. The parameter θ_2 describes the correlation among the variables. When $\theta_2 > 0$ the interactions among agents tends to align the value of relational variables and, we may have that almost no pair of agents are connected or either almost all pair of agents are connected.

Result 18 suggests that conclusions based on hypothesis testing with ERGM are not valid when analysing a single network unless additional auxiliary assumptions on the network formation process are imposed. In statistical social network analysis, multimodal distributions have often been rejected on the argument that for a single network data, unimodal distributions have to be used to fit the data (Snijders, 2002; Snijders et al., 2006). It has been added as an assumption by stating that we observe the expected network statistics (Cranmer and Desmarais, 2011); and bimodal distributions have been suggested as an indication of a misspecified model (Li, 2015). Further, bimodal distributions have also been classified as near-degeneracy (Robins et al., 2007), where a model is said to be near-degenerate when it places almost all its probability mass around a small subset of networks (e.g. the null network, the complete network or both of them) (Handcock et al., 2003). Another example of the assumption of unimodal distribution is presented in a generalisation of the ERGM termed *statistical exponential random graph models* (SERGM) (Chandrasekhar and Jackson, 2014). For the SERGM, conditions for consistency are binding to assume that the network statistics concentrate around their expected value (see Supplementary Materials).

Briefly, the rationale for rejecting models with multimodal distributions is that these models are likely to place negligible mass on the observed statistics (which equals the expected value of the estimated model, $\mathrm{E}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\Gamma})$). On the other hand, when ERGM have unimodal distributions, then 1) the network models based on potential games admit only one class of Nash Equilibria or agents' decisions are made almost surely at random, and 2) any finite collection of relational variables are asymptotically independent (Bhamidi et al., 2011).

In regression models, the unimodal distribution of statistics are a consequence of the independence assumption and the law of large numbers. Researchers can control the validity of the assumption by controlling the data-generating process (Freedman, 1987). For network models, it is an ad hoc assumption needed to justify statistical inference.

Two solutions have been proposed to the issue of explicitly stating the assumptions that validate law of large numbers and central limit theorem for the network statistics. First, replace assumptions on large networks by assumptions on large samples (several observed networks). This approach is appealing as it can be applied in studies when several networks have been collected. However, a strongest limitation of this approach is that it cannot be applied to the large class of empirical studies with a single observed network.

A second solution, like the one presented in Section 6.2, is to look for conditions under which the test statistics converge to a known distribution, even when the relational variables are dependent. Previous work in this direction was taken by Schweinberger and Handcock (2015). Schweinberger and Handcock (2015) approach characterised the structural dependency by imposing dependency between local variables. As we will see in Section 6.7, a limitation, though, is that auxiliary assumptions that validate hypothesis testing on a single network are sufficient for treating a single network as several independent hyponetworks.

6.6.3 The community fallacy

Although, as previously discussed, bimodal distribution have previously been claimed to be caused by misspecified models. This type of distributions have a long history in the mathematical framework surrounding ERGM and they have been observed in several social systems. Therefore, it is important to understand under which conditions bimodal distribution can be observed in ERGM. In this subsection, we show that bimodal distribution naturally occur in the mathematical framework used in ERGM. Then, we show that the distribution of the network statistics at the community level and at the macroscopic level are quite different. On the one hand, social mechanisms with opposing forces may drive the system at the community level to have statistics with bimodal distributions. On the other hand, the distribution at the macroscopic level, may be far from being close to the null network or the complete network. Thus, inference at the community level cannot be generalised to the macroscopic-level, and vice versa.

6.6.4 Bimodal distribution occurring in complex systems

When the network statistics defining an ERGM stems from assumptions on dependency structure between variables (as it is done in Assumption 6), we have that the ERGM is grounded on the probabilistic framework of Markov random fields, i.e. a collection of variables and an undirected network describing the dependency between the variables. Historically, Markov random fields were a mathematical construction for putting the Ising model into a probabilistic setting (Kindermann et al., 1980). This model was proposed by Wilhelm Lenz and Ernest Ising in the 1920's, and it is a mathematical model for explaining empirical facts about ferromagnetism.

It is defined by a graph or a lattice with N nodes and associated with each node i there is a spin variable σ_i taking only the values -1 and $+1$. Nearby spins influenced each other, and each spin is influenced by an external magnetic field. The probability of observing a particular realisation of the spins is given by

$$P_\beta(\boldsymbol{\sigma}) = \frac{\exp\big(-\beta H(\boldsymbol{\sigma})\big)}{Z(\beta)}$$

$$H(\boldsymbol{\sigma}) = -J\sum_{<i,j>}\sigma_i\sigma_j - H\sum_{i=1}^{N}\sigma_i$$

where $<i,j>$ indicates the pair of linked nodes in the graph or lattice. $H(\boldsymbol{\sigma})$ is known as the Hamiltonian (energy) of the system for the realisation $\boldsymbol{\sigma}$. The parameter H, J are called the magnetic field and the coupling constant, respectively. $\beta = \frac{1}{T}$ is called the inverse temperature, T the temperature and $Z(\beta)$ is a normalising constant. As observed by Park and Newman (2004), the two-star model can be thought as an Ising model on the edge-dual graph of the complete graph. The connection between the Ising model and potential games was observed by Auletta et al. (2016).

The property of interest is the total magnetisation $M = \frac{1}{N}\sum_{i=1}^{N}\sigma_i$, and it ranges from -1 to $+1$. It was believed that for sufficiently low temperatures, the interactions between the spins would cause them to exhibit ferromagnetism, that is, most spins would tend to align to the value either -1 or $+1$. In the 1920's, Ernest Ising proved that there was no ferromagnetic behaviour in the one-dimensional lattice at any temperature; and he wrongly conjectured that a similar result would occur in higher dimensions. In 1936, Rudolf Peierls proved that the model does display spontaneous magnetisation at low temperatures in the two-dimensional lattice (Peierls, 1936). Peierls can be regarded as the first to notice the mathematical similarity between the Ising model and cooperative phenomena (Niss, 2005), i.e. phenomena in which the elements of a system cannot be considered as acting independently from each other.

Beyond its applications in ERGM, the Ising model can be seen as a simple model for social

In the dual-edge graph of a graph, the edges are replaced by nodes and the nodes are replaced by edges.

interactions where each agent tend to mimic the actions of its neighbours (Stauffer, 2008). For instance, as a simple opinion dynamics model in which the opinion of an agent is influenced by the majority of their interacting partners, the Ising model shows that interactions may result in the system being in two possible ordered states (e.g. positive or negative opinion) for low temperatures, while above a critical temperature the system remains macroscopically disordered (Castellano et al., 2009; Galam and Martins, 2015). Market events have been studied as herd behaviour caused by traders mimicking the behaviour of others, and bubbles and crashes are originated by local self-reinforcing of these imitations (Johansen et al., 2000). Business confidence has been shown to be mostly in states of "collective pessimism" or "collective optimism", and this bimodality has been modelled using the Ising model by Hohnisch et al. (2006).

In summary, bimodal distributions is not an isolated phenomena of ERGM, but they are the result of cooperative phenomena.

6.6.5 Distribution at the community and macroscopic level

The next two examples show that bimodality can be explained as the observation of the relationship between agents in only one community. For the sake of simplicity, we will focus on the undirected case. The first model is defined by 105 agents grouped in 15 communities, and each community has 7 agents. The network statistics defining the model is the number of relations, the number of relations between agents in the same community and the number of two stars between agents in the same community and their respective parameters are -2.4, 1 and 0.55. In Figure 17 (A), we observe that the distribution of the number of relations between agents in the same community is bimodal with several agents being connected with almost no other agent in their respective community while others are connected with almost all agents in their respective community. In Figure 18, we see that the network statistics (number of relations- and the number of two-stars between agents in the same community) are clustered in two regions: one around the null network and another around the complete network. Further, the distribution places most of the probability on network statistics close to the boundary of the convex hull. On the other hand, the distribution of the average of the network statistics over the 15 communities is not longer bimodal, but it is unimodal, and the distribution places most of the probability on averages not close to the boundary of the convex hull. In Figure 17 (B), we observe that the number of relations in the network is distributed around 160. Thus, generated networks at random from this model are not likely to be the null network or the complete network. The average transitivity is around 0.53, the average diameter of a network 12.52 and the average number of isolated agents is approximately 15%.

An interpretation of this model is that almost all agents within a community are related to each other for long periods of time but then abruptly almost all relations between them disappear, which is the result of the self-reinforced behaviour to remove relations. It stays like this for a long period, and suddenly almost all agents in a community are again related to each other. For long periods, a large proportion of agents remains connected via relations with agents in other communities.

Our second model is defined by 70 agents grouped in 10 communities; each community has 7 agents. The network statistics defining the model is the number of relations, the number of relations between agents in the same community and the number of two stars-, number of triangles between agents in the same community and their respective parameters are -1.2, 0.5, -0.1 and 1.2. As with the previous model, we have that the distribution of the number of relations between agents in the same community is bimodal while the total number of relations is unimodal, see Figure 19. The network statistics defined by relations between agents in the same community are again clustered in two regions: one around the null network and another around the complete network, see Figure 20. However, the distribution of the average of the network statistics over the 10 communities is not longer bimodal, but it is unimodal, and the distribution does not place most of its probability to network statistics near to the network statistics of the null network or complete network.

Simulated networks from the model show that the average transitivity is around 0.17, the average diameter of a network is around 4.3 and the average number of isolated agents is less than 1%.

The previous results show that network properties at the community level can be quite different from network properties at the macroscopic level, and vice-versa. In particular, the undesirable property of having networks with almost everyone connected or disconnected is not inherited when it occurs at the community level. This is important, since discerning if a bimodal distribution or an estimation problem is caused by a misspecified model or a small sample problem is crucial for choosing the remedy. If the problem is diagnosed as a misspecified model, it implies for the researcher removing and introducing variables to the initial model. However, if the real problem is a small sample, the remedy is likely to cause specification bias.

6.7 Analysing multiple networks or a single large network?

Although, the methods for analysing a single network have been claimed to be more challenging than analysing a single network. In Section 6.7.1, we show that estimators of one most popular methods for analysing multiple networks do not have fundamental statistical properties, and thus conclusions based on this approach are not reliable. Next, in section 6.7.2, we show that a recent method for analysing large networks treats a large network as several independent networks.

6.7.1 The analysis of multiple networks

In various settings of social interactions, it is known that agents are nested within groups. In educational research many problems have a hierarchical structure, students are nested within classrooms, classrooms are nested within schools (De Leeuw et al., 2008). For instance, students' educational performance needs to incorporate a hierarchical structure to its analysis, as it is claimed that students' performance is the outcome of agents' characteristics, i.e. family background, and school-level characteristics, i.e. classroom size, teacher's experience (Darling-Hammond, 2000; Hill et al., 2005). When network data possess a hierarchical structure, statistical network models need to incorporate information about the agents' characteristics, the dependency structure between the relational variables and additionally the groups' characteristics (Vermeij et al., 2009; Bellotti, 2012). The statistical models for analysing hierarchical network data were coined Multilevel network models (Lazega and Snijders, 2015). Two problems proposed in social networks with a hierarchical structure are 1) whether the composition of classes influences the formation process of network and 2) to what extent the differences in network structures between classes is caused by random noise or the composition of classes (Lubbers, 2003).

Although theoretically, Multilevel network models require sampling several observed networks while only one observed network is required for C-ERGM; they are closely connected via the concept of groups in Multilevel network models and the concept of communities presented in C-ERGM. First, in multilevel network models, it is also assumed that the updating processes of relations between groups are independent. Second, as the number of observed groups increases in Multilevel network models, estimators are expected to be more accurate (Lazega and Snijders, 2015). For the sake of simplicity, from now on we will replace the term group by community.

A sharp difference between Multilevel network models and C-ERGM is the use of the information of the communities. For C-ERGM, the analysis of formation of networks is only possible by observing a *large* network, where large is not longer in terms of the number of agents but on the number of communities in the networks. On the other hand, one of the most common estimation method used in Multilevel network models (i.e. two-step procedures) performs statistical analysis on each observed network as if a network consists of multiple observations. However, estimation problems observed in empirical studies using two-step regression methods (Lubbers, 2003; Kruse et al., 2016), is explained by not treating networks with only one community as single observations (Guardian et al., 2016). Next, we present in detail the inconsistencies of two-step procedures first documented in (Guardian et al., 2016).

In Multilevel network models, the formation of networks is broken into two levels: (1) a micro-level describing the updating process of the relations between agents within a community and (2) a macro-level describing the influence the community's characteristics has on the updating process in each community. It is assumed that the updating process of relations within a particular community is independent of the updating processes occurring in others groups. Modelling the dependencies between the relational variables in multilevel network analysis has been done using ERGM in the micro-level. While the effects of the community characteristics are determined by a set of linear relations between ERGM-parameters and some community-level variables (Lubbers, 2003).

In its simplest form, the two-step procedures assume that networks $\{\boldsymbol{e_n}(i)\}_{i=1}^{\infty}$ are independent random graphs but not identically distributed with $\boldsymbol{e_n}(i) \sim P_{\boldsymbol{\theta}(i)}$. $\{P_{\boldsymbol{\theta}(i)}\}_{i=1}^{\infty}$ are ERGM defined by the same network statistics but with different unknown parameters. The parameter $\boldsymbol{\theta}$ is assumed to be a random variable,

$$\boldsymbol{\theta} = \boldsymbol{\mu_\theta} + \boldsymbol{\epsilon} \tag{79}$$

with $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$.

The estimator procedure is then divided into two steps. In the first step, a maximum likelihood estimator $\hat{\boldsymbol{\theta}}(i)$ is fitted for each network $\boldsymbol{e_n}(i)$. In the second step, the unknown parameter in Formula (79) is replaced by its estimator.

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\mu_\theta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta} \tag{80}$$

with $\boldsymbol{\zeta}$ equals the random variable $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$. The simplest estimator for $\boldsymbol{\mu_\theta}$ for a sample of m networks is the average of the estimators $\hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = \frac{1}{m}\sum_{i=1}^{m} \hat{\boldsymbol{\theta}}(i)$ (Snijders and Baerveldt, 2003). For the sake of simplicity, let us focus on the k^{th} parameter.

Hypothesis testing on the parameter μ_{θ_k} is done by constructing a t-statistic and assuming that it is normally distributed (Lubbers, 2003; Lubbers and Snijders, 2007; Valente et al., 2009;

If the data is temporal, then the micro-level can be modelled with an stochastic actor oriented model (Snijders and Baerveldt, 2003)

In practice, the parameter is estimated using an MCMC-MLE.

Schaefer et al., 2011; Huitsing and Veenstra, 2012; McFarland et al., 2014; An, 2015; Boda and Néray, 2015; Kruse et al., 2016). The validity of the hypothesis testing leans on the following points:

1. $\hat{\mu}_{\theta_k}$ converge in probability to the real μ_{θ_k}

2. $\frac{(\hat{\mu}_{\theta_k}-\mu_{\theta_k})}{\sqrt{v\hat{a}r(\hat{\mu}_{\theta_k})}}$ converge to a normal distribution $\mathcal{N}(0,1)$

$v\hat{a}r(\hat{\mu}_{\theta_k})$ is an estimator for the variance of $\hat{\mu}_{\theta_k}$.

For point (1) and (2) to hold, there are two components of the model where assumptions can be imposed. One is to makes assumptions on the error term ϵ_k and the second is to introduce assumptions on the ERGM. The last point is rarely discussed, despite the fact that for any class of ERGM there always exists certain networks such that log-likelihood function does not have maximum. A common solution to this problem is to remove from the analysis networks where the log-likelihood function does not have a maximum or when the algorithm does not converge. This solution, though, requires the removal of an infinite number of networks in the limit when m tends to infinity.

If the MLE is defined as the argument that maximise the log likelihood function when the observed network lies in $Conv^{\circ}(\mathcal{C}_n)$ and it assigns an arbitrary value otherwise, then the estimator $\hat{\mu}_{\theta_k}$ converge in probability to $\mu_0^k + bias_{\theta_k}$.

$$bias_{\theta_k} = E(\zeta_{k,0}) = \int_{\{\mathbf{\Gamma}(\boldsymbol{e_n})\in Conv^{\circ}(\mathcal{C}_n)\}}(\hat{\theta}_k(\boldsymbol{e_n})-\mu_0^k)dP_{\boldsymbol{\theta}}(\boldsymbol{e_n})+\int_{\{\mathbf{\Gamma}(\boldsymbol{e_n})\in\partial Conv(\mathcal{C}_n)\}}(\hat{\theta}_k(\boldsymbol{e_n})-\mu_0^k)dP_{\boldsymbol{\theta}}(\boldsymbol{e_n})$$

The assumption of unbiased, i.e. $bias_{\theta_k} = 0$, is likely to be an inaccurate approximation of reality when the sizes of the communities under analysis are small, e.g. $P_{\boldsymbol{\theta}}(\{\mathbf{\Gamma}(\boldsymbol{e_n}) \in \partial Conv(\mathcal{C}_n)\}) \not\approx 0$ (as MLE takes an arbitrary value) or $\mathrm{E}(\hat{\theta}_k|\{\mathbf{\Gamma}(\boldsymbol{e_n}) \in Conv^{\circ}(\mathcal{C}_n)\})$ is not approximately equal to μ_0^k. However, when the rationale to justify unbiased estimators is observing large networks, it shifts the biased problem to the consistency problem in ERGM.

6.7.2 Limitations of methodologies for analysing large networks

The former observations raise the question till what extend are statistical tools for analysing several networks different from the ones used for analysing large networks? Schweinberger and Handcock (2015) claimed that the study of large networks is challenger than statistical analysis of multiple networks, and they provide theoretical justifications of statistical inference on large network by proposing a local dependence condition. Broadly speaking, the local dependence condition states: (a) e_{ij} and e_{kl} are dependent if agents i, j, k and l belong to the same community but (b) a relational variable e_{ij} between agents in different communities is independent of all other variables except possible for e_{ji}.

Our analysis of the local dependence condition will be done by showing that C-ERGM satisfies a slightly weaker local dependence condition by point (b) is relaxed. Our next result shows that assumptions 5, 6 and 7 divide the relational variables into disjoint independent blocks. Each block can be decomposed in two type of ERGMs. The first ones are similar to the classical ERGM which define a probability space over networks, while the others are ERGM defining a probability space over relational variables between a pair of communities (hyponetworks).

Result 19. *Let $P_{\boldsymbol{\theta}^n}$ be a sequence of ERGM satisfying Assumptions 5-7 with partition $\{U_i\}_{i=1}^{\infty}$. Let $\{U_i^n\}_{i=1}^{n_u}$ denote the partition of the agents $V(\boldsymbol{e_n})$ in communities by $U_i^n = U_{j_i} \bigcap V(\boldsymbol{e_n})$ where $\{U_{j_i}\}_{i=1}^{n_u}$ denotes all communities in the partition such that $U_{j_i} \bigcap V(\boldsymbol{e_n}) \neq \emptyset$.*

Let $\mathcal{U}_i^n$ be the set of pair of agents in U_i^n, and $\{W_i\}_{i=1}^{n_w}$ are the sets of pairs of agents between two communities $\{U_i^n\}_{i=1}^{n_u}$.

The probability distribution $P_{\boldsymbol{\theta}^n}$ equals

$$P_{\boldsymbol{\theta}^n}(\boldsymbol{e_n}) = \prod_{i=1}^{n_u} P_{\boldsymbol{\theta}_{\boldsymbol{u}}^n}(\boldsymbol{e}_{\mathcal{U}_i^n}) \prod_{i=1}^{n_w} P_{\boldsymbol{\theta}_{\boldsymbol{w}}^n}(\boldsymbol{e}_{W_i}) \tag{81}$$

where $\boldsymbol{\theta}_{\boldsymbol{u}}^{\boldsymbol{n}} = (\theta_1, ..., \theta_p)$, $\boldsymbol{\theta}_{\boldsymbol{w}}^{\boldsymbol{n}} = (\theta_{p+1}, ..., \theta_{p+q})$. and

$$P_{\boldsymbol{\theta}_{\boldsymbol{u}}^{n}}(\boldsymbol{e}_{\mathcal{U}_i^n}) = \frac{\exp\left(\sum_{k=1}^{p}\theta_k^n\Gamma_k(\boldsymbol{e}_{\mathcal{U}_i^n})\right)}{\sum_{\bar{\boldsymbol{e}}_{\mathcal{U}_i^n}}\exp\left(\sum_{k=1}^{p}\theta_k^n\Gamma_k(\boldsymbol{e}_{\mathcal{U}_i^n})\right)}$$

and

$$P_{\boldsymbol{\theta}_{\boldsymbol{w}}^{n}}(\boldsymbol{e}_{W_i^n}) = \frac{\exp\left(\sum_{k=p+1}^{p+q}\theta_k^n\Gamma_k(\boldsymbol{e}_{W_i})\right)}{\sum_{\bar{\boldsymbol{e}}_{W_i}}\exp\left(\sum_{k=p+1}^{p+q}\theta_k^n\Gamma_k(\boldsymbol{e}_{W_i})\right)}.$$

A goal of introducing the local dependence condition was to show a Gaussian behaviour of the network statistics defined by aggregating information of relations within communities and between communities.

$$\frac{\Gamma_{k,U}(\boldsymbol{e_n}) + \Gamma_{k,W}(\boldsymbol{e_n}) - \bar{\Gamma}_k}{\sqrt{\mathrm{Var}}(\Gamma_{k,U}(\boldsymbol{e_n}) + \Gamma_{k,W}(\boldsymbol{e_n}))} \to \frac{\Gamma_{k,U}(\boldsymbol{e_n}) - \bar{\Gamma}_k}{\sqrt{\mathrm{Var}}(\Gamma_{k,U}(\boldsymbol{e}_n))} \to \mathcal{N}(0,1) \tag{82}$$

where $\bar{\Gamma}_k$ is the expected network statistic under the true model, $\mathrm{E}\big(\Gamma_{k,U}(\boldsymbol{e_n}) + \Gamma_{k,W}(\boldsymbol{e_n})\big)$.

Additionally from the local dependence condition, the convergence in distribution in (82) requires that networks are δ-sparse. A random graph is said to be δ-sparse if there exists a constant c and a $\delta > 0$ such that

$$\mathrm{E}(|e_{ij}|^p) \leq cn^{-\delta} \quad p = 1,2$$

for all relational variables e_{ij} with agents i, j belonging to different communities. Sparsity aims to express the notion that agents cannot maintain an arbitrary number of relations, but it has the dire consequence that under middle conditions, the expected number of relations between agents in different communities converge to zero, e.g. for large networks the communities are almost completely isolated (see Supplementary Materials):

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}^n}(\Gamma_{1,W}(\boldsymbol{e_n}) > \epsilon) = 0 \ \ \forall\epsilon > 0$$

A different characterisation of sparsity was presented in (Chandrasekhar and Jackson, 2016).

An example of the limitation of sparsity is that it cannot express the idea of disassortative, as it would imply that the expected number of relations converges to zero. Sparsity is a requisite since network data is summarised by aggregating the information of the within communities and the between communities, but the information of between communities has to be negligible. The work of Schweinberger and Handcock (2015) can be seen as requirements needed for performing statistical inference when it is known the existence of a community structure satisfying certain conditions but without knowing to which community each agent belongs to, so that all relational information have to be aggregated. Here, the difference between Multi-level network models and C-ERGM on the one hand, and the one proposed by Schweinberger and Handcock (2015), on the other, is that in the last one, there is missing information about the community structure; but both break the generating process of networks into independent processes.

By reducing the updating process of the relations occurring in a set of agents to completely independent updating processes, we can model each process in a community with an ERGM. The next result justifies the use of the maximum likelihood estimator (MLE) by taking communities as the units of observation. It also provides arguments in favour that the unit of observation for ERGM is the community an not the agents or the relational variables.

Result 20. *Let Assumptions 5-8 be satisfied and let* $\{U_i^n\}_{i=1}^{n_u}$, $\{W_i\}_i^{n_w}$ *be as in Result 19 with the cardinality of* U_i^n *equals* B *for all* i *and large* n*, and the parameter be a constant function of the number of agents,* $\boldsymbol{\theta}$.

If we denote the maximum likelihood estimators as

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{\boldsymbol{n}} = \arg\max_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \sum_{i=1}^{n_l} \Big(\sum_{k=1}^{p} \tilde{\theta}_k \Gamma_k(\boldsymbol{e}_{\mathcal{U}_i}) - \log(c_n(\tilde{\boldsymbol{\theta}})) \Big)$$

and $\mathrm{Cov}_{\boldsymbol{\theta}_{\boldsymbol{u}}}(\boldsymbol{\Gamma}_U)$ *is positive definite, then*

1.

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{\boldsymbol{n}} \to \boldsymbol{\theta}_{\boldsymbol{u}}$$

2. If we denote the maximum likelihood estimators as

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}^{\boldsymbol{n}} = \arg\max_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^q} \sum_{i=1}^{n_w} \Big(\sum_{k=p+1}^{p+q} \tilde{\theta}_k \Gamma_k(\boldsymbol{e}_{W_i}) - c_n(\tilde{\boldsymbol{\theta}}) \Big)$$

and $\mathrm{Cov}_{\boldsymbol{\theta}_{\boldsymbol{w}}}(\boldsymbol{\Gamma}_W)$ *is positive definite; then*

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}^{\boldsymbol{n}} \to \boldsymbol{\theta}_{\boldsymbol{w}}$$

and

$$\frac{\Gamma_k(\boldsymbol{e}_{\boldsymbol{n}}) - \mathrm{E}_{\boldsymbol{\theta}}(\Gamma_k)}{\sqrt{\mathrm{Var}_{\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}}(\Gamma_k)}} \xrightarrow{D} \mathcal{N}(0,1) \tag{83}$$

The difference between the Limit (83) and the limit of Theorem 9 is that we replace the variance of the network statistics by the estimated variance.

The limit in Formula (82) also has the true variance and not an estimated variance.

6.8 Discussion

Network studies are moving from exploratory analyses to testing hypothesis about the social mechanisms underlying the formation of networks. However, hypothesis testing puts the burden of proof on the plausibility of the model assumptions on the defenders of the model. One of the most important models for hypothesis testing on a single network is the exponential random graph model (ERGM) . The importance of model assumptions for ERGM was shown by the fact that p-values on a single network are validated under the conjecture that the test statistics are asymptotically normal (Hunter and Handcock, 2006), but little is known on the assumptions under which the conjecture holds. In this paper, we showed via mod-ϕ convergence (Féray et al., 2016) the existence of a subclass of ERGM (C-ERGM) for which asymptotic normality of the test statics holds. C-ERGM assumes the existence and knowledge of disjoint communities of the agents, and it can be seen as a mixed between ERGM and stochastic blockmodels. Despite, their appealing for hypothesis testing some problems persists: (1) C-ERGM partitions the relational variables into disjoint blocks and (2) C-ERGM assumes independence between variables in different blocks. Solutions to Points (1) and (2) require either a priori definition of the communities and the existence of this information in the collected data or the definition of the communities via a community detection algorithm (Fortunato, 2010).

Our paper also raises problems on the rationale behind p-values for ERGM based on having networks with a large number of agents and not on having several communities in the network. We show that introducing community structure to ERGM helps to understand that estimation problems are not necessarily caused by misspecified models, as it is commonly assumed, but they are also the result of small sample size problem. Further, methods for analysing multiple networks fail to solve the small sample size problem, as first documented by (Guardian et al., 2016). The acknowledge small sample size problems in network studies is indispensable for not overstating the strength of the results and overfitting. Noble Prize winner Clive W.J Granger said: "If you can't get it right as n goes to infinity, you shouldn't be in this business". When studying a single network, Granger's remark becomes particularly important as the complexity of network data impose auxiliary assumptions for the construction of statistical evidence that is seldom justified or discussed.

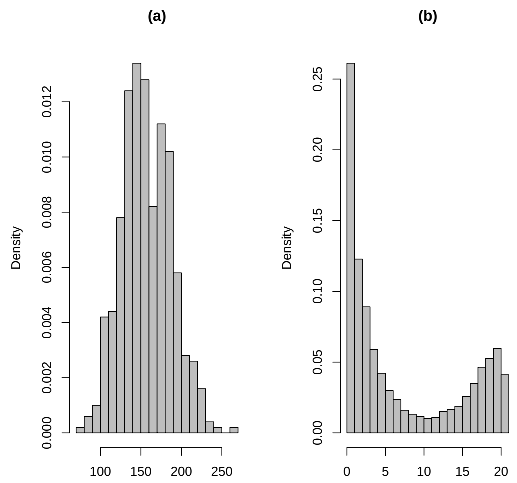


Figure 17: Histogram **(a)** shows that the distribution of the number of relations at the macroscopic level is unimodal; while histogram **(b)** shows that the distribution of the number of relations at the community level is bimodal.

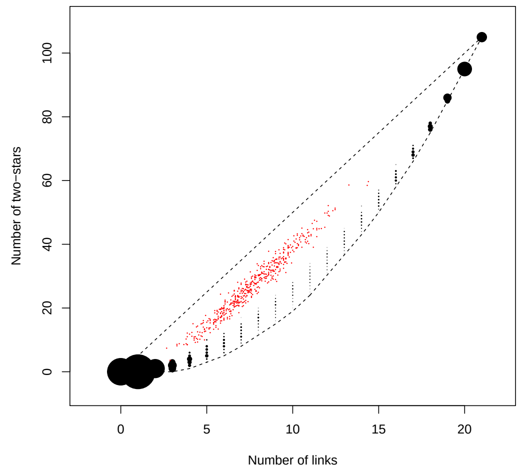


Figure 18: The joint distribution of the network statistics number of relations and number of two-stars within a community is represented with black points, and the size is proportional to the frequency of occurrence. Most of the probability mass is around the boundary of the convex hull, and this implies estimation problems. The red points represent the distribution of the network statistics for the complete network. In this case, most of the probability mass is in the interior of the convex hull.

References

Albert, A. and Anderson, J. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71(1):1–10.

An, W. (2015). Multilevel meta network analysis with application to studying network dynamics of network interventions. *Social Networks*, 43:48–56.

Anderson, C. J., Wasserman, S., and Faust, K. (1992). Building stochastic blockmodels. *Social networks*, 14(1):137–161.

Auletta, V., Ferraioli, D., Pasquale, F., Penna, P., and Persiano, G. (2016). Convergence to equilibrium of logit dynamics for strategic games. *Algorithmica*, 76(1):110–142.

Babichenko, Y. and Tamuz, O. (2016). Graphical potential games. *Journal of Economic Theory*, 163:889–899.

Barndorff-Nielsen, O. (2014). *Information and exponential families in statistical theory*. John Wiley & Sons.

Bellotti, E. (2012). Getting funded. multi-level network of physicists in italy. *Social Networks*, 34(2):215–229.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.

Bhamidi, S., Bresler, G., and Sly, A. (2011). Mixing time of exponential random graphs. *Annals of Applied Probability*, 21(6):2146–2170.

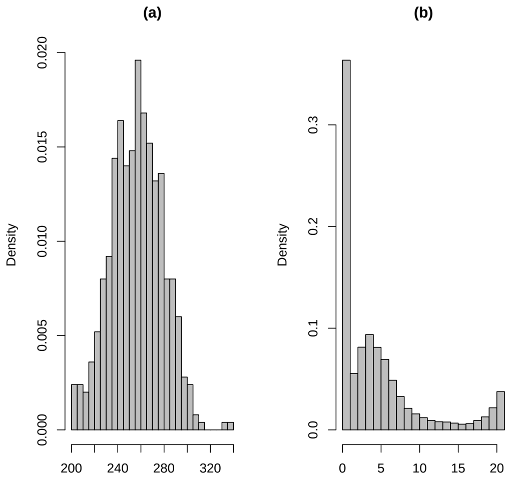


Figure 19: Histogram**(a)** shows that the distribution of the number of relations at the macroscopic level is unimodal; while histogram **(b)** shows that distribution of the number of relations at the community level is bimodal.

Boda, Z. and Néray, B. (2015). Inter-ethnic friendship and negative ties in secondary school. *Social Networks*, 43:57–72.

Byshkin, M., Stivala, A., Mira, A., Krause, R., Robins, G., and Lomi, A. (2016). Auxiliary parameter mcmc for exponential random graph models. *Journal of Statistical Physics*, 165(4):740–754.

Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591.

Chandrasekhar, A. G. and Jackson, M. O. (2014). Tractable and consistent random graph models. Technical report, National Bureau of Economic Research.

Chandrasekhar, A. G. and Jackson, M. O. (2016). A network formation model based on subgraphs.

Collaboration, O. S. et al. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716.

Cox, D. (1982). Statistical significance tests. *British journal of clinical pharmacology*, 14(3):325–331.

Cranmer, S. J. and Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Political Analysis*, 19(1):66–86.

Cranmer, S. J., Desmarais, B. A., and Menninga, E. J. (2012). Complex dependencies in the alliance network. *Conflict Management and Peace Science*, 29(3):279–313.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education policy analysis archives*, 8:1.

De Leeuw, J., Meijer, E., and Goldstein, H. (2008). Handbook of multilevel analysis.

Eichelsbacher, P., Martschink, B., et al. (2015). On rates of convergence in the Curie–Weiss–Potts model with an external field. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 51, pages 252–282. Institut Henri Poincaré.

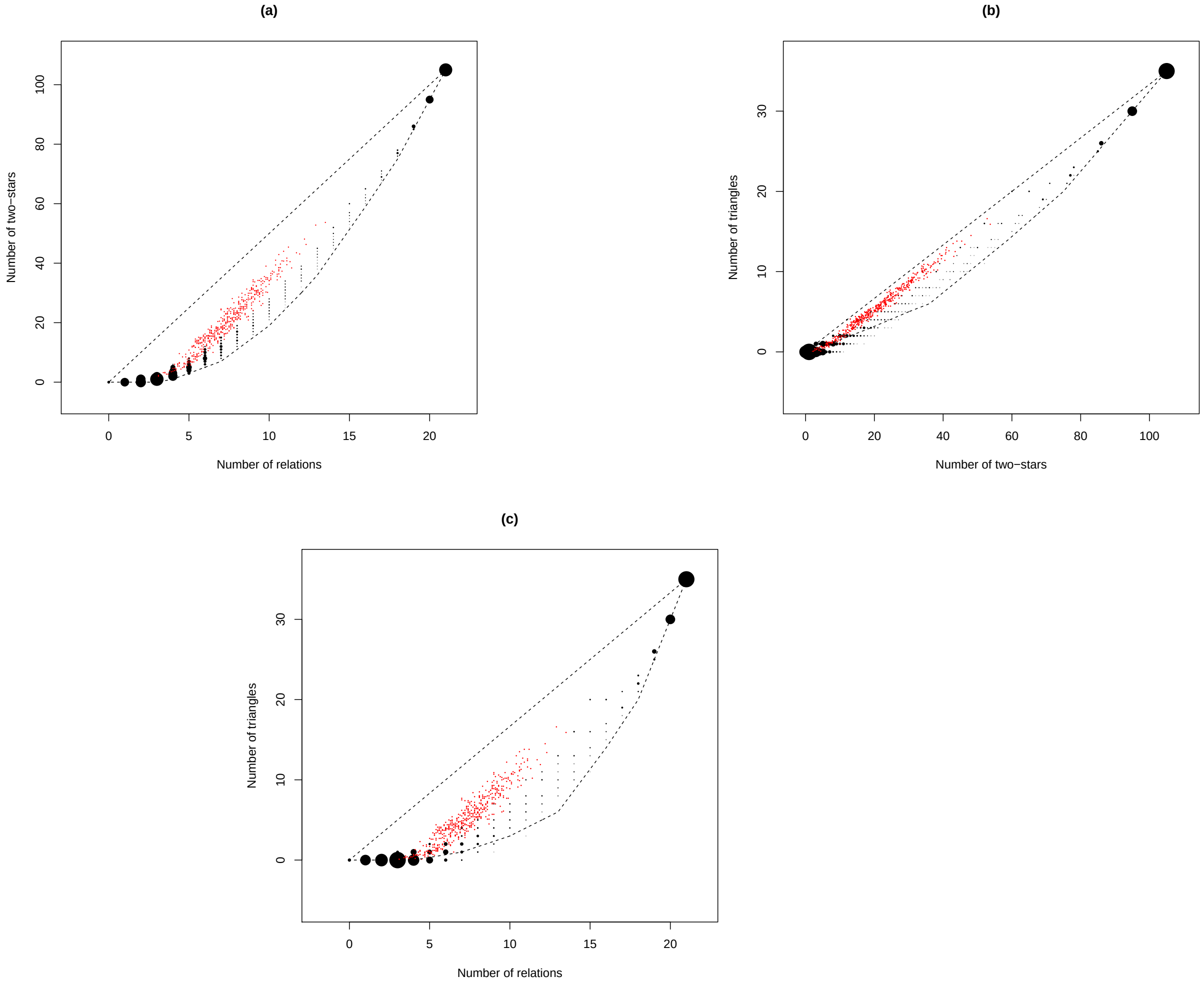


Figure 20: **(a)**-**(c)** shows the probability distribution of the network statistics, the black points represent the possible values of the network statistics within communities while the red points represent the possible values of the network statistics for the complete network. The size of the points is proportional to the frequency of occurrence.

Ellis, R. S. and Newman, C. M. (1978). Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 44(2):117–139.

Ellis, R. S., Newman, C. M., and Rosen, J. S. (1980). Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 51(2):153–169.

Ellwardt, L., Wittek, R., and Wielers, R. (2012). Talking about the boss: Effects of generalized and interpersonal trust on workplace gossip. *Group & organization management*, page 1059601112450607.

Féray, V., Méliot, P., and Nikeghbali, A. (2015a). Mod-φ convergence, ii: Estimates of the speed of convergence and local limit theorems.

Féray, V., MÉLIOT, P.-L., and NIKEGHBALI, A. (2015b). Mod-φ convergence, i: Normality zones and precise deviations. *arXiv preprint arXiv:1304.2934*.

Féray, V., Meliot, P.-L., and Nikeghbali, A. (2016). *Mod-φ convergence: Normality zones and precise deviations*. Springer.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368.

Fisher, R. A. (1935). The logic of inductive inference. *Journal of the royal statistical society*, 98(1):39–82.

Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3):75–174.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.

Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational and Behavioral Statistics*, 12(2):101–128.

Galam, S. and Martins, A. C. (2015). Two-dimensional Ising transition through a technique from two-state opinion-dynamics models. *Physical Review E*, 91(1):012108.

Gandolfo, D., Ruiz, J., and Wouts, M. (2010). Limit theorems and coexistence probabilities for the Curie–Weiss Potts model with an external field. *Stochastic Processes and their Applications*, 120(1):84–104.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33(5):587–606.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233.

Guardian, A. C., Tessone, C. J., and Algesheimer, R. (2016). The effects the number of agents has on the formation of networks and statistical analysis on multiple networks.

Häggström, O. and Jonasson, J. (1999). Phase transition in the random triangle model. *Journal of Applied Probability*, pages 1101–1115.

Handcock, M. S., Robins, G., Snijders, T., Moody, J., and Besag, J. (2003). Assessing degeneracy in statistical models of social networks. In *Journal of the American Statistical Association*.

Hanneke, S., Fu, W., Xing, E. P., et al. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605.

Hayashi, F. (2000). Econometrics.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., and Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biol*, 13(3):e1002106.

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in medicine*, 21(16):2409–2419.

Hill, H. C., Rowan, B., and Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American educational research journal*, 42(2):371–406.

Hohnisch, M., Stauffer, D., and Pittnauer, S. (2006). The impact of external events on the emergence of social herding of economic sentiment. *arXiv preprint physics/0606237*.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

Huitsing, G. and Veenstra, R. (2012). Bullying in classrooms: Participant roles from a social network perspective. *Aggressive Behavior*, 38(6):494–509.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3).

Ilany, A., Barocas, A., Koren, L., Kam, M., and Geffen, E. (2013). Structural balance in the social networks of a wild mammal. *Animal Behaviour*, 85(6):1397–1405.

Jacobsen, M. (1989). Existence and unicity of mles in discrete exponential family distributions. *Scandinavian Journal of Statistics*, pages 335–349.

Johansen, A., Ledoit, O., and Sornette, D. (2000). Crashes as critical points. *International Journal of Theoretical and Applied Finance*, 3(02):219–255.

Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107.

Kindermann, R., Snell, J. L., et al. (1980). *Markov random fields and their applications*, volume 1. American Mathematical Society Providence.

Kruse, H., Smith, S., van Tubergen, F., and Maas, I. (2016). From neighbors to school friends? how adolescents? place of residence relates to same-ethnic school friendships. *Social Networks*, 44:130–142.

Lazega, E. and Snijders, T. A. (2015). Multilevel network analysis for the social sciences.

Lewis, K. (2013). The limits of racial prejudice. *Proceedings of the National Academy of Sciences*, 110(47):18814–18819.

Li, K. (2015). *Degeneracy, Duration, and Co-evolution: Extending Exponential Random Graph Models (ERGM) for Social Network Analysis*. PhD thesis.

Lorenz, G. and Hall, R. L. (2013). The biases of binary data: Partisan and revolving door ties in a count-edge network of health reform lobbyists. In *APSA 2013 Annual Meeting Paper*.

Lubbers, M. J. (2003). Group composition and network structure in school classes: a multilevel application of the p* model. *Social Networks*, 25(4):309–332.

Lubbers, M. J. and Snijders, T. A. (2007). A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social Networks*, 29(4):489–507.

Lubell, M., Scholz, J., Berardo, R., and Robins, G. (2012). Testing policy theory with statistical models of networks. *Policy Studies Journal*, 40(3):351–374.

Lusher, D., Koskinen, J., and Robins, G. (2012). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press.

Lusher, D., Kremer, P., and Robins, G. (2013). Cooperative and competitive structures of trust relations in teams. *Small group research*, page 1046496413510362.

McFarland, D. A., Moody, J., Diehl, D., Smith, J. A., and Thomas, R. J. (2014). Network ecology and adolescent social structure. *American sociological review*, page 0003122414554001.

Meester, R., Collins, M., Gill, R., and Van Lambalgen, M. (2006). On the (ab) use of statistics in the legal case against the nurse lucia de b. *Law, Probability and Risk*, 5(3-4):233–250.

Mele, A. (2013). A structural model of segregation in social networks. *Available at SSRN 2294957*.

Méliot, P.-L. and Nikeghbali, A. (2015). Mod-gaussian convergence and its applications for models of statistical mechanics. In *In Memoriam Marc Yor-Séminaire de Probabilités XLVII*, pages 369–425. Springer.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Neyman, J. and Pearson, E. S. (1992). On the problem of the most efficient tests of statistical hypotheses. In *Breakthroughs in Statistics*, pages 73–108. Springer.

Niss, M. (2005). History of the lenz-ising model 1920–1950: from ferromagnetic to cooperative phenomena. *Archive for history of exact sciences*, 59(3):267–318.

O'Malley, A. J. (2013). The analysis of social network data: an exciting frontier for statisticians. *Statistics in medicine*, 32(4):539–555.

Park, J. and Newman, M. E. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6):066117.

Pattison, P. and Robins, G. (2002). Neighborhood-based models for social networks. *Sociological Methodology*, 32(1):301–337.

Peierls, R. (1936). On Ising's model of ferromagnetism. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 32, pages 477–481. Cambridge Univ Press.

Ricciuti, R. (2008). Bringing macroeconomics into the lab. *Journal of Macroeconomics*, 30(1):216–237.

Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social networks*, 29(2):192–215.

Schaefer, D. R., Simpkins, S. D., Vest, A. E., and Price, C. D. (2011). The contribution of extracurricular activities to adolescent friendships: New insights through social network analysis. *Developmental Psychology*, 47(4):1141.

Schweinberger, M. and Handcock, M. S. (2015). Local dependence in random graph models: characterization, properties and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):647–676.

Siciliano, M. D. (2015). Advice networks in public organizations: The role of structure, internal competition, and individual attributes. *Public Administration Review*.

Smith, V. L. (2002). Method in experiment: Rhetoric and reality. *Experimental economics*, 5(2):91–110.

Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1):361–395.

Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models.

Snijders, T. A. and Baerveldt, C. (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of Mathematical Sociology*, 27(2-3):123–151.

Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153.

Song, H. (2015). Uncovering the structural underpinnings of political discussion networks: Evidence from an exponential random graph model. *Journal of Communication*, 65(1):146–169.

Stauffer, D. (2008). Social applications of two-dimensional Ising models. *American Journal of Physics*, 76(4):470–473.

Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204–212.

Tatarenko, T. (2014). Proving convergence of log-linear learning in potential games. In *American Control Conference (ACC), 2014*, pages 972–977. IEEE.

Uddin, S., Hamra, J., and Hossain, L. (2013). Exploring communication networks to understand organizational crisis using exponential random graph models. *Computational and Mathematical Organization Theory*, 19(1):25–41.

Valente, T. W., Fujimoto, K., Chou, C.-P., and Spruijt-Metz, D. (2009). Adolescent affiliations and adiposity: a social network analysis of friendships and obesity. *Journal of Adolescent Health*, 45(2):202–204.

Vermeij, L., Van Duijn, M. A., and Baerveldt, C. (2009). Ethnic segregation in context: Social discrimination among native dutch pupils and their ethnic minority classmates. *Social Networks*, 31(4):230–239.

Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to Markov graphs. *Psychometrika*, 61(3):401–425.

Wasserstein, R. L. and Lazar, N. A. (2016). The asa's statement on p-values: context, process, and purpose. *The American Statistician*, (just-accepted):00–00.

Wu, B., Jiang, S., and Chen, H. (2015). The impact of individual attributes on knowledge diffusion in web forums. *Quality & Quantity*, 49(6):2221–2236.

7 Supplementary Materials (C)

7.1 Exponential random graph model with community structure

Theorem 8 is an immediate consequence of the more general Theorem 11.

Let us define a function $f : \mathbb{B} \to D \subset \mathbb{N}$. We say that that two hyponetworks $\bar{\boldsymbol{e}}_S$ and $\bar{\boldsymbol{e}}_{S'}$ are f-isomorphic ($\bar{\boldsymbol{e}}_S \overset{f}{\sim} \bar{\boldsymbol{e}}_{S'}$) if there exists a bijective function $\varphi : V(\boldsymbol{e}_S) \to V(\boldsymbol{e}_{S'})$ such that 1) for all $e_{ij} \in \boldsymbol{e}_S$ $\bar{e}_{ij} = \bar{e}_{\varphi(i)\varphi(j)}$ and for all $e_{i'j'} \in \boldsymbol{e}_{S'}$ $\bar{e}_{i'j'} = \bar{e}_{\varphi^{-1}(i')\varphi^{-1}(j')}$. 2) for all $i \in V(\boldsymbol{e}_S)$, we have that $f(i) = f(\varphi(i))$. As before, we define a ***structure*** as a set of hyponetworks isomorphic to each other, and when a hyponetwork is in a structure we say that the hyponetwork has that structure.

For instance, if $D = \{0, 1\}$ and 1 represents female while 0 represents men, then f-isomorphism preserves a) number of females and males and b) the relations between agents of the same (different) gender.

Point 2) can also be replaced by other constraints. An important case is 2') for all $e_{ij} \in \boldsymbol{e}_S$ $f(i) - f(j) = f(\varphi(i)) - f(\varphi(j))$. When for all $i, j \in V(\boldsymbol{e}_S)$ there exists a sequence $\{i_1, i_2, ..., i_n\}$ with $i_1 = i$, $i_n = j$ and $e_{i_w i_{w+1}} \in \boldsymbol{e}_S$; we have that if $\boldsymbol{e}_S$ and $\boldsymbol{e}_{S'}$ are f-isomorphic with bijective function φ, for all $i \in V(\boldsymbol{e}_S)$ $f(\varphi(i)) = f(i) + c$ for some constant c.

The reason are twofold. On the one hand, we have that for any agent i, there exist a constant c with $f(\varphi(i)) = f(i) + c$. Second, for any agents $i, j \in V(\boldsymbol{e}_S)$, we have that $f(\varphi(j)) - f(\varphi(i)) = f(j) - f(i)$ as for there exists a sequence $\{i_1, i_2, ..., i_n\}$ with $i_1 = i$, $i_n = j$ and $e_{i_w i_{w+1}}$ such that

$$f(\varphi(j)) - f(\varphi(i)) = \sum_{w=1}^{n-1} f(\varphi(i_{w+1})) - f(\varphi(i_w)) = \sum_{w=1}^{n-1} f(i_{w+1}) - f(i_w) = f(j) - f(i).$$

Thereby, we have that $f(\varphi(j)) = f(j) + c$

When $f(i)$ represents the community in which agent i belongs, f-isomorphism preserves the number of communities and the existence and no-existence of relations between and within communities.

Theorem 11. *Let us assume that the random variables $\boldsymbol{e}_S$ satisfy the following conditions*

For all $e_{ij} \in \boldsymbol{e}_S$, we have that

$$P(e_{ij}|e_{-ij}) = c_{ij} \prod_{k=1}^{p} \prod_{l=1}^{l_k} \prod_{\substack{\boldsymbol{e}_S \in S_k \\ e_{ij} \in \boldsymbol{e}_S}} Q_S^{(k,l)}(\boldsymbol{e}_S). \tag{84}$$

for some c_{ij}.

Functions $Q_S^{(k,l)}(\boldsymbol{e}_S)$ are of the form

$$Q_S^{(k,l)}(\boldsymbol{e}_S) = \begin{cases} c_{(k,l)} & \textit{if } \boldsymbol{e}_S \overset{f_l}{\sim} H_k \\ 1 - c_{(k,l)} & \textit{otherwise} \end{cases} \tag{85}$$

If all realisations have positive probability, then the joint distribution of $\boldsymbol{e}_S$ equals

$$P_{\boldsymbol{\theta}}(\boldsymbol{e}_S) = \frac{1}{c(\boldsymbol{\theta})} \exp\left(\sum_{k=1}^{p} \sum_{l=1}^{l_k} \theta_{k,l} \Gamma_{k,l}(\boldsymbol{e}_S) \right). \tag{86}$$

where $\theta_{k,l} = \log\left(\frac{c_{(k,l)}}{1-c_{(k,l)}}\right)$ and $\Gamma_{k,l}(\boldsymbol{e}_S)$ is the number of subnetworks in $\boldsymbol{e}_S$ f_l-isomorphic to H_k. $c(\boldsymbol{\theta})$ is the normalising constant.

Proof of Theorem 11:

The probability distribution (86) equals

$$P_{\boldsymbol{\theta}}(\boldsymbol{e}_S) = \frac{1}{c(\boldsymbol{\theta})} \exp\Big(\sum_{k=1}^{p} \sum_{l=1}^{l_k} \sum_{\substack{\boldsymbol{e}_S \in S_k \\ e_{ij} \in \boldsymbol{e}_S}} \theta_{k,l} \mathbb{1}_{\{\boldsymbol{e}_S \overset{f_l}{\sim} H_k\}} \Big) \tag{87}$$

Recall that $P_{\boldsymbol{\theta}}(e_{ij}|e_{-ij})$ equals $\frac{P_{\boldsymbol{\theta}}(e_{ij},e_{-ij})}{P_{\boldsymbol{\theta}}(e_{ij}=1,e_{-ij})+P_{\boldsymbol{\theta}}(e_{ij}=0,e_{-ij})}$. The numerator and denominator are equal to the following two expressions

$$\overbrace{\frac{1}{c(\boldsymbol{\theta})}\exp\Big(\sum_{k=1}^{p}\sum_{l=1}^{l_k}\sum_{\substack{\boldsymbol{e}_S\in S_k\\ e_{ij}\notin\boldsymbol{e}_S}}\theta_{k,l}\mathbb{1}_{\{\boldsymbol{e}_S\overset{f_l}{\sim}H_k\}}\Big)}^{A}\exp\Big(\sum_{k=1}^{p}\sum_{l=1}^{l_k}\sum_{\substack{\boldsymbol{e}_S\in S_k\\ e_{ij}\in\boldsymbol{e}_S}}\theta_{k,l}\mathbb{1}_{\{\boldsymbol{e}_S\overset{f_l}{\sim}H_k\}}\Big) \tag{88}$$

$$\begin{aligned}&\overbrace{\frac{1}{c(\boldsymbol{\theta})}\exp\Big(\sum_{k=1}^{p}\sum_{l=1}^{l_k}\sum_{\substack{\boldsymbol{e}_S\in S_k\\ e_{ij}\notin\boldsymbol{e}_S}}\theta_{k,l}\mathbb{1}_{\{\boldsymbol{e}_S\overset{f_l}{\sim}H_k\}}\Big)}^{A}\times\\ &\Bigg(\overbrace{\exp\Big(\sum_{k=1}^{p}\sum_{l=1}^{l_k}\sum_{\substack{\boldsymbol{e}_S\in S_k\\ e_{ij}\in\boldsymbol{e}_S\\ e_{ij}=1}}\theta_{k,l}\mathbb{1}_{\{\boldsymbol{e}_S\overset{f_l}{\sim}H_k\}}\Big)+\exp\Big(\sum_{k=1}^{p}\sum_{l=1}^{l_k}\sum_{\substack{\boldsymbol{e}_S\in S_k\\ e_{ij}\in\boldsymbol{e}_S\\ e_{ij}=0}}\theta_{k,l}\mathbb{1}_{\{\boldsymbol{e}_S\overset{f_l}{\sim}H_k\}}\Big)}^{m_{ij}}\Bigg)\end{aligned} \tag{89}$$

By noticing that m_{ij} is a constant function of e_{ij} and combining Formula (88) and (89), we get that

$$P(e_{ij}|e_{-ij})=d_{ij}\exp\Big(\sum_{k=1}^{p}\sum_{l=1}^{l_k}\sum_{\substack{\boldsymbol{e}_S\in S_k\\ e_{ij}\in\boldsymbol{e}_S}}\theta_{k,l}\mathbb{1}_{\{\boldsymbol{e}_S\overset{f_l}{\sim}H_k\}}\Big) \tag{90}$$

where d_{ij} equals $\frac{1}{m_{ij}}$.

Finally, observer that $Q_S^{(k,l)}=c\exp\big(\log(\frac{c_{(k,l)}}{1-c_{(k,l)}})\mathbb{1}_{\{\boldsymbol{e}_S\overset{f_l}{\sim}H_k\}}\big)$. where c is a constant.

This holds, since

$$Q_S^{(k,l)}=(1-c_{(k,l)})\Big(\frac{c_{(k,l)}}{1-c_{(k,l)}}\Big)^{\mathbb{1}_{\{\boldsymbol{e}_S\overset{f_l}{\sim}H_k\}}}=\exp\Big(\log(\frac{c_{(k,l)}}{1-c_{(k,l)}})\mathbb{1}_{\{\boldsymbol{e}_S\overset{f_l}{\sim}H_k\}}\Big)\exp\left(1-c_{(k,l)}\right)$$

■

Theorem 8 follows from Theorem 11 by taking f the function that maps each agent to its community. For the functions $Q_{i,k}$ $(i = 1, 2)$ in Assumption 6, we replace the isomorphism by $\overset{f}{\sim}$. For $i = 1$, we add to H_k the assumption that all agents belong to the same community. While for $i = 2$, we add to H_{k_r} the assumption that agents can be partition in two communities such that for $e_{ij} \in V(H_{k_r})$, i, j are not in the same community.

Proof of Result 8:

Follows from Theorem 11.

■

The assumptions of Theorem 11 defines a Markov random field and it is possible to describe the joint distribution in terms of the cliques in its dependency graph, but using Formulas (85), the joint distribution can be represented in terms of the network statistics. This representation, though, in term of network statistics is not necessarily unique.

The proof of Theorem 9 leans on the concept of mod-ϕ convergence. Mod-ϕ convergence has been shown to be a powerful tool for obtaining central limit theorems for dependent variables Méliot and Nikeghbali (2015) and other limits distributions, as the one due to Ellis et al. (1980).

Definition 13. *(Féray et al. (2015a)) Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real valued function and D a subset of the complex numbers containing 0. We suppose there is a non-constant infinitely divisible distribution ϕ with $\int_{\mathbb{R}} e^{zx}\phi(x)dx = e^{\eta(z)}$. The sequence $\{X_n\}_{n\in\mathbb{N}}$ is said to converge mod-ϕ over D with parameter t_n and limiting function ψ, if locally uniformly on D,*

$$\mathbb{E}(e^{zX_n})e^{\frac{-t_n\eta(z)}{2}} \to \psi(z)$$

where ψ is a continuous function on D with $\psi(0) = 1$ and $t_n \to \infty$. When $D = i\mathbb{R}$, we simply say mod-ϕ convergence.

When ϕ is the standard Gaussian distribution, we speak of mod-Gaussian convergence. It was observed in Méliot and Nikeghbali (2015) that if a sequence is mod-Gaussian, then the central limit theorem holds, i.e. $\frac{X_n}{\sqrt{t_n}}$ converges in distribution to $\mathcal{N}(0,1)$.

A simple proof of a slightly weaker version of Theorem 9 is possible using the results presented by Féray et al. (2015b) using Theorem 12.

Theorem 12. *(Féray et al., 2015b) Let $\{Y_l\}_1^{N_n}$ be a set of variables and D_n the maximum degree of their dependency graph. Let us define $X_n = \sum_{l=1}^{N_n} Y_l$ and let us assume that the following conditions holds*

1. $\lim_{n\to\infty} \frac{D_n}{N_n} = 0$.

2. *There exists a positive constant C such that for all $r \geq 2$, $|K^r_{\frac{X_n}{D_n}}(0)| \leq (Cr)^r \frac{N_n}{D_n}$.*

3.
$$\sigma_n^2 = \frac{D_n}{N_n} K^{(2)}_{\frac{X_n}{D_n}}(0) \to \sigma > 0$$
$$L_n = \sigma_n^3 \frac{D_n}{N_n} K^{(3)}_{\frac{X_n}{D_n}}(0) \to L$$

where $K_X(t)$ denotes the cumulant generating function for the random variable X, $K_X(t) = \log(E(e^{tX}))$ and its k-th derivative is denoted by $K_X^{(r)}(t)$.

Then
$$\frac{X_n - E(X_n)}{D_n \sigma_n} \Big(\frac{D_n}{N_n}\Big)^{\frac{1}{3}}$$

converges in the mod-Gaussian sense with parameter $t_n = \left(\frac{N_n}{D_n}\right)^{\frac{1}{3}}$ and limiting function $\psi(z) = \exp\left(\frac{Lz^3}{6}\right)$.

For $k \in \{1, ..., p\}$, we show that $\Gamma_k^n = \Gamma_k(\boldsymbol{e_n})$ is the sum of variables such that the maximum degree of their dependency graph satisfies point 1 of Theorem 12. For this, let us assume without loss of generality that $V(H_k)$ has cardinality h_k; and let $\mathcal{U}_{n,k}$ be the set of subsets of $\mathcal{F}_n$ such that if $e_S \in \mathcal{U}_{n,k}$ then the cardinality of $V(\boldsymbol{e}_S)$ is exactly h_k. For any $S \in \mathcal{U}_{n,k}$, we define the indicator variables

$$\mathbb{1}_k(\boldsymbol{e}_S) = \begin{cases} 1 & \text{if } \boldsymbol{e}_S \sim H_k \text{ and } V(e_S) \subset U_l \text{ for some l} \\ 0 & \text{otherwise} \end{cases}$$

First, notice that by construction that Γ_k^n is equal to the sum $\sum_{S \in \mathcal{U}_{n,k}} \mathbb{1}_k(\boldsymbol{e}_S)$. We construct the dependency graph $\boldsymbol{G}_{n,k}$ for the indicator variables in $\mathcal{U}_{n,k}$ as follows. The nodes are the above-constructed indicator functions. Two nodes $\mathbb{1}_k(\boldsymbol{e}_S)$, $\mathbb{1}_k(\boldsymbol{e}_{S'})$ are connected if the two variables are dependent. They are dependent when at least one of the following two conditions is satisfied: (1) $\boldsymbol{e}_S \cap \boldsymbol{e}_{S'} \neq \emptyset$ or (2) there exists a variable in $\boldsymbol{e}_S$ that is dependent on a variable in $\boldsymbol{e}_{S'}$. Finally, the next result shows that Point 1 in Theorem 12 is satisfied when Assumption 5 holds.

Result 21. *If $D_{n,k}$ denotes the maximum degree and n_k denotes the number of nodes of the dependency graph $\boldsymbol{G}_{n,k}$, then*

$$\lim_{n \to \infty} \frac{D_{n,k}}{n_k} = 0.$$

Proof of Result 21:

The proof is based on two observation.

First, notice that a node $\mathbb{1}_k(\boldsymbol{e}_S)$ in $\boldsymbol{G}_{n,k}$ is connected to at most $(2B)_k = 2B(2B - 1)...(2B - k + 1)$ other nodes when there exist at most two communities $U_l, U_{l'}$ such that $V(\boldsymbol{e}_S) \subseteq U_l \bigcup U_{l'}$. If set $V(\boldsymbol{e}_S)$ is not contained in the union of two communities, then $\mathbb{1}_k(\boldsymbol{e}_S)$ is the constant function and it is independent of all other indicator variables, which means that $\mathbb{1}_k(\boldsymbol{e}_S)$ is an isolated node in $\boldsymbol{G}_{n,k}$.

Second,the nodes of the graph $\boldsymbol{G}_{n,k}$ is by construction the cardinality of $\mathcal{U}_{n,k}$, which has cardinality equal to the falling factorial $(n)_k = n(n-1)...(n-k+1)$.

Thereby, the maximum degree is $o((n)_k)$.

■

Next, we prove a simple version of our main theorem assuming that the parameters are constant and replacing Assumption 8 by the stronger assumption:

Assumption 18.

$$\sigma_n^2 = \frac{\mathrm{Var}_{\boldsymbol{\theta}^n}(\Gamma_k)}{n_k} \to \sigma > 0$$

$$L_n = \sigma_n^3 \frac{D_{n,k}}{n_k} K^{(3)}_{\frac{\Gamma_k}{D_{n,k}}}(0) \to L$$

Proof of weaker version of Theorem 9:

Point 1 follows from Result 21.

Point 2 follows from the upper bound $|K^{(r)}_{\Gamma_k^n}(0)| \leq (2^{r-1}r^{r-2})(D_{n,k}+1)^{r-1}n_k$ established by Féray et al. (2015b). In particular,

$$|K^{(r)}_{\Gamma_k^n}(0)| \leq (2r)^r (D_{n,k})^{r-1} n_k.$$

Using the property that for a positive constant a, $K^{(r)}_{\frac{\Gamma_k^n}{a}}(t) = \frac{1}{a^r} K^{(r)}_{\Gamma_k^n}(t)$ in the last equation, we have

$$\left|K^{(r)}_{\frac{\Gamma_k^n}{D_{n,k}}}(0)\right| \leq (2r)^r \frac{n_k}{D_{n,k}}.$$

Point 3 follows from Assumption 18.

As a consequence, the assumptions of Theorem 12 are valid and we have that $\frac{\Gamma_k^n - E(\Gamma^n)}{\sigma_n(k)D_{n,k}} \left(\frac{D_{n,k}}{n_k}\right)^{\frac{1}{3}}$ converge in the mod-Gaussian sense with $t_n = \left(\frac{n_k}{D_{n,k}}\right)^{\frac{1}{3}}$ and limiting function $\psi(z) = \exp\left(\frac{Lz^3}{6}\right)$. Thereby,

$$\frac{\Gamma_k^n - E(\Gamma_k^n)}{\sigma_n(k)D_{n,k}} \left(\frac{D_{n,k}}{n_k}\right)^{\frac{1}{3}} \left(\frac{n_k}{D_{n,k}}\right)^{\frac{1}{6}} \to \mathcal{N}(0,1)$$

Rearranging terms in the last equation, we have that

$$\frac{\Gamma_k^n - \mathrm{E}_{\boldsymbol{\theta}^n}(\Gamma_k^n)}{\sqrt{\mathrm{Var}_{\boldsymbol{\theta}^n}(\Gamma_k^n)}} \xrightarrow{D} \mathcal{N}(0,1)$$

■

Bounding the size of the communities by a constant in Assumption 5 is only used in the above proof via Result 21.

The proof of Theorem 10 leans on the next Theorem presented in Féray et al. (2015a).

Theorem 13. *(Féray et al., 2015a) Let* $S_n = \sum_{i=1}^{N_n} Y_{i,n}$ *be a sequence of sums of centred dependent random variables, all bounded in absolute value by* 1*, and with dependency graph* G_n *of parameters* $N_n \to \infty$ *and* $D_n = o(N_n)$*. We suppose* $Y_{i,n}$ *for all* i, n *and*

$$\lim_{n\to\infty} \inf \frac{\mathrm{Var}(S_n)}{N_n D_n} = \sigma^2 > 0. \tag{91}$$

Then, for n *large enough*

$$d_K\Big(\frac{S_n}{\sqrt{\mathrm{Var}(S_n)}}, \mathcal{N}(0,1)\Big) \leq \frac{56}{\sigma^3}\sqrt{\frac{D_n}{N_n}}$$

Proof of Theorem 10: For a k, let us define $Y_{S,n} = \mathbb{1}_k(\boldsymbol{e}_S) - \mathrm{E}\big(\mathbb{1}_k(\boldsymbol{e}_S)\big)$. By definition $Y_{S,n}$ is centred and its absolute value is smaller or equal to one. Further, $S_n = \sum_{S\in\mathcal{U}_{n,k}} \Big(\mathbb{1}_k(\boldsymbol{e}_S) - \mathrm{E}\big(\mathbb{1}_k(\boldsymbol{e}_S)\big)\Big) = \Gamma_k^n - E(\Gamma_k^n)$ and $\mathrm{Var}(S_n) = \mathrm{Var}(\Gamma_k^n)$.

In particular, we have

$$\frac{S_n}{\sqrt{\mathrm{Var}(S_n)}} = \frac{\Gamma_k^n - E(\Gamma_k^n)}{\sqrt{\mathrm{Var}(\Gamma_k^n)}}.$$

Hence, it is sufficient to show that Theorem 13 holds for the variables $Y_{S,n}$.

By construction, the dependency graph defined by the centred variables $Y_{S,n}$ is isomorphic to the dependency graph defined by the indicator variables $\mathbb{1}_k(\boldsymbol{e}_S)$. Thus, the dependency graph of the variables $Y_{S,n}$ satisfies that $D_n = o(N_n)$.

Limit (91) is satisfied by Assumption 8.

■

7.2 Network models and data-assumptions

Proof of Result 11:

Equation (74) follows from the Hammersley-Clifford theorem (Besag, 1974).

Let us consider the partition $\overset{B}{\sim}$ induced by the set of variables $\boldsymbol{e}_B$, let $[\boldsymbol{e}_{\mathcal{C}}]$ denotes a class, and let us assume that e_{ij} and e_{kl} are conditionally independent on the realisation $\bar{\boldsymbol{e}}_B$ with $e_{ij}, e_{kl} \in \boldsymbol{e}_C \setminus \boldsymbol{e}_B$.

By observing that $\boldsymbol{e}_{S'} \overset{B}{\sim} \boldsymbol{e}_{S''}$ if and only if $\boldsymbol{e}_{S'} \setminus \boldsymbol{e}_B = \boldsymbol{e}_{S''} \setminus \boldsymbol{e}_B$, we have that

$$\Big(\prod_{e_{i'j'} \in \boldsymbol{e}_C \setminus \boldsymbol{e}_B} e_{ij}\Big)\Big(\sum_{\boldsymbol{e}_{S'} \in [\boldsymbol{e}_C]} \theta_{S'} \prod_{e_{i'j'} \in \boldsymbol{e}_{S'} \cap \boldsymbol{e}_B} e_{i'j'}\Big) = \sum_{\boldsymbol{e}_{S'} \in [\boldsymbol{e}_C]} \theta_{S'} \prod_{e_{i'j'} \in \boldsymbol{e}_{S'}} e_{i'j'}$$

The constraints of Equation (75) follow from the fact that the equality $P(e_{ij}, e_{kl} | \boldsymbol{e}_B = \bar{\boldsymbol{e}}_B) = P(e_{ij} | \boldsymbol{e}_B = \bar{\boldsymbol{e}}_B) P(e_{kl} | \boldsymbol{e}_B = \bar{\boldsymbol{e}}_B)$ implies

$$\Big(\prod_{e_{i'j'} \in \boldsymbol{e}_C \setminus \boldsymbol{e}_B} e_{ij}\Big)\Big(\sum_{\boldsymbol{e}_{S'} \in [\boldsymbol{e}_C]} \theta_{S'} \prod_{e_{i'j'} \in \boldsymbol{e}_{S'} \cap \boldsymbol{e}_B} \bar{e}_{i'j'}\Big) = 0$$

Since, $\prod_{e_{i'j'} \in \boldsymbol{e}_C \setminus \boldsymbol{e}_B} e_{ij}\Big) \neq 0$, we have that

$$\sum_{\boldsymbol{e}_{S'} \in [\boldsymbol{e}_C]} \theta_{S'} \prod_{e_{i'j'} \in \boldsymbol{e}_{S'} \cap \boldsymbol{e}_B} \bar{e}_{i'j'} = 0$$

■

Proof of Result 12:

If $\boldsymbol{e}_B = \bar{\boldsymbol{0}}$ and $\boldsymbol{e}_{S'} \cap \boldsymbol{e}_B \neq \emptyset$, then $\prod_{e_{i'j'} \in \boldsymbol{e}_{S'} \cap \boldsymbol{e}_B} \bar{e}_{i'j'} = 0$.

As a result,

$$\theta_{S''} = \sum_{\boldsymbol{e}_{S'} \in [\boldsymbol{e}_C]} \theta_{S'} \big(\prod_{e_{i'j'} \in \boldsymbol{e}_{S'} \cap \boldsymbol{e}_B} \bar{e}_{i'j'}\big) = 0$$

where $\boldsymbol{e}_{S''}$ is the unique element in $[\boldsymbol{e}_C]$ such that $\boldsymbol{e}_{S''} \cap \boldsymbol{e}_B = \emptyset$. ■

Proof of Result 13:

If we define $\bar{\boldsymbol{e}} = (\bar{e}_i, \bar{\boldsymbol{e}}_{-i})$ and $\tilde{\boldsymbol{e}} = (\tilde{e}_i, \bar{\boldsymbol{e}}_{-i})$, then

$$\Phi(\bar{\boldsymbol{e}}) - \Phi(\tilde{\boldsymbol{e}}) = \sum_{k=1}^{p} \theta_k \Gamma_k(\bar{\boldsymbol{e}}) - \sum_{k=1}^{p} \theta_k \Gamma_k(\tilde{\boldsymbol{e}}) \overset{(a)}{=}$$

$$\sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'}\in S_k}\theta_k\mathbb{1}_{\left\{\substack{i\in V(\bar{\boldsymbol{e}}_{S'})\\ \bar{\boldsymbol{e}}_{S'}\sim H_k}\right\}}+\sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'}\in S_k}\theta_k\mathbb{1}_{\left\{\substack{i\notin V(\bar{\boldsymbol{e}}_{S'})\\ \bar{\boldsymbol{e}}_{S'}\sim H_k}\right\}}-\sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'}\in S_k}\theta_k\mathbb{1}_{\left\{\substack{i\in V(\tilde{\boldsymbol{e}}_{S'})\\ \tilde{\boldsymbol{e}}_{S'}\sim H_k}\right\}}-\sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'}\in S_k}\theta_k\mathbb{1}_{\left\{\substack{i\notin V(\tilde{\boldsymbol{e}}_{S'})\\ \tilde{\boldsymbol{e}}_{S'}\sim H_k}\right\}}\overset{(b)}{=}$$

$$\sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'}\in S_k}\theta_k\mathbb{1}_{\left\{\substack{i\in V(\bar{\boldsymbol{e}}_{S'})\\ \bar{\boldsymbol{e}}_{S'}\sim H_k}\right\}}-\sum_{k=1}^{p}\sum_{\boldsymbol{e}_{S'}\in S_k}\theta_k\mathbb{1}_{\left\{\substack{i\in V(\tilde{\boldsymbol{e}}_{S'})\\ \tilde{\boldsymbol{e}}_{S'}\sim H_k}\right\}}=u_i(\bar{\boldsymbol{e}}_{S'})-u_i(\tilde{\boldsymbol{e}}_{S'})$$

Equality (a) follows by definition, and equality (b) follows by the fact that $\bar{\boldsymbol{e}}_{S'}=\tilde{\boldsymbol{e}}_{\boldsymbol{S'}}$ if $i\notin V(\boldsymbol{e}_{S'})$.

Proof of Result 14:

From the previous result, we have that

$$\Phi(\bar{\boldsymbol{e}})>\Phi(\tilde{\boldsymbol{e}}) \tag{92}$$

if and only if

$$u_i(\bar{\boldsymbol{e}})>u_i(\tilde{\boldsymbol{e}}) \tag{93}$$

If we multiply by a positive number, β, both Sides of the Equations (92) and (93) and we apply a monotonic function, $\exp()$, the inequalities are preserved. Thereby,

$$P_{\boldsymbol{\theta}^{\beta}}(\bar{\boldsymbol{e}})>P_{\boldsymbol{\theta}^{\beta}}(\tilde{\boldsymbol{e}})$$

if and only if

$$v_i^{\beta}(\bar{\boldsymbol{e}})>v_i^{\beta}(\tilde{\boldsymbol{e}})$$

■

The proof for the Result 15 follows from Theorem 4 (Tatarenko, 2014).

Proofs of Result 16 and Result 17:

It is not difficult to observe that the constructed Markov chain is aperiodic and irreducible. Since the number of states is finite, the construed Markov chain is positive recurrent. As a result, there exists a unique stationary distribution, which equals the limiting distribution, π^{∞}.

$$\lim_{\tau \to \infty} \boldsymbol{Q}\pi = \pi^{\infty}$$

where $\boldsymbol{Q}$ is the transition matrix and π is any initial distribution.

In order to prove that $\pi^{\infty} = P_{\boldsymbol{\theta}}$, it is sufficient to show that $\sum_{\bar{\boldsymbol{e}}_n \in \mathcal{G}^n} Q_{\tilde{\boldsymbol{e}},\bar{\boldsymbol{e}}} P_{\boldsymbol{\theta}}(\bar{\boldsymbol{e}}) = P_{\boldsymbol{\theta}}(\tilde{\boldsymbol{e}})$.

$$\sum_{\bar{\boldsymbol{e}}_n \in \mathcal{G}^n} Q_{\tilde{\boldsymbol{e}},\bar{\boldsymbol{e}}} P_{\boldsymbol{\theta}}(\bar{\boldsymbol{e}}) = \sum_{\bar{\boldsymbol{e}}_n \in \mathcal{G}^n} \Big(\frac{Q_{\bar{\boldsymbol{e}},\tilde{\boldsymbol{e}}} P_{\boldsymbol{\theta}}(\tilde{\boldsymbol{e}})}{P_{\boldsymbol{\theta}}(\bar{\boldsymbol{e}})}\Big) P_{\boldsymbol{\theta}}(\bar{\boldsymbol{e}}) = P_{\boldsymbol{\theta}}(\tilde{\boldsymbol{e}}) \underbrace{\sum_{\bar{\boldsymbol{e}}_n \in \mathcal{G}^n} Q_{\bar{\boldsymbol{e}},\tilde{\boldsymbol{e}}}}_{1} = P_{\boldsymbol{\theta}}(\tilde{\boldsymbol{e}})$$

■

7.3 Counter example of the central limit theorem

Proof of Result 18:

We study the properties of the constructed ERGM by observing that it is a special case of the Curie-Weiss Potts model (Ellis and Newman, 1978; Gandolfo et al., 2010; Eichelsbacher et al., 2015).

The joint probability distribution of the defined class of ERGM equals

$$P_{(h,\beta,n)}(\boldsymbol{e_m}) = \frac{1}{c_n(\boldsymbol{\theta^n})} \exp\Big(\theta_1 \sum_{e_{ij} \in \boldsymbol{e_n}} e_{ij} + \frac{\theta_2}{2m_n} \sum_{e_{ij} \neq e_{i'j'}} e_{ij} e_{i'j'}\Big)\Big). \tag{94}$$

Next, we do a change of variable by defining the set of symmetric variables $\boldsymbol{x_n} \in \{-\frac{1}{2}, \frac{1}{2}\}^m$ where $x_{ij} = \frac{1}{2} - e_{ij}$, $h = \theta_1$ and $\beta = \theta_2$.

$$P_{(h,\beta,n)}(\boldsymbol{x_n}) = \frac{1}{Z(h,\beta,n)} \exp\Big(-h_n \sum_{x_{ij}\in \boldsymbol{x_n}} x_{ij} + \frac{\beta}{2m_n} \sum_{x_{ij}\neq x_{i'j'}} x_{ij}x_{i'j'}\Big). \tag{95}$$

where $Z(h,\beta,n)$ is the normalizing constant and $h_n = (h-\beta)(1+O(\frac{1}{m_n}))$.

Model (95) is the Curie-Weiss Potts model for 2-states and $m_n = n(n-1)$ spins (Eichelsbacher et al., 2015).

If we denote $S_n(\boldsymbol{x_n}) = \sum_{e_{ij}\in \boldsymbol{e_n}} e_{ij}$, then $S_n(\boldsymbol{x_n}) = \Gamma_1(\boldsymbol{e_n}) - \frac{m_n}{2}$ and we know its asymptotic distribution (Theorem 2 Ellis and Newman (1978)).

Case 1. If 1.1) $h-\beta > 0$ and $\beta \geq 0$ or 1.2) $h-\beta = 0$ and $\beta \in (0,4)$, then $\frac{S_{m_n}}{\sqrt{m_n}}$ converges to the normal distribution

$$\mathcal{N}\big(0, \sigma^2(\beta)\big).$$

Case 2. $h-\beta = 0$ and $\beta = 4$, $m_n^{-\frac{3}{4}} S_{m_n}$ converges to the distribution F_2 where F_2 has probability density $f_2(x) \propto \exp(-\frac{4}{3}t^4)$.

Case 3. $h-\beta = 0$ and $\beta > 4$, then $\frac{S_{m_n}}{m_n}$ converges in distribution to $\frac{1}{2}\delta_{u_1(\beta)} + \frac{1}{2}\delta_{u_2(\beta)}$ with $u_1(\beta)$, $u_2(\beta)$ being the two global minimum of the Function (96).

For the cases 1.1), 2) and 3), the asymptotic behaviour of S_m depends on the extremal points of the following function:

$$G_\beta(s) = \frac{\beta s^2}{2} - \log\big(\cosh(\frac{\beta s}{2})\big). \tag{96}$$

When $\beta \leq 4$, there is a unique global minimum $s=0$ and $\frac{S_{m_n}}{m_n}$ converges to 0; when $\beta > 4$ the function has two global minimum ($u_1(\beta)$, $u_2(\beta)$) and an inflection point 0. This is shown by noticing that $G'_\beta(s)$ has a unique root ($s=0$) when $\beta \leq 4$ and three roots otherwise.

$$G'_\beta(s) = \beta s - \frac{\beta}{2}\tanh\Big(\frac{\beta s}{2}\Big). \tag{97}$$

When $\beta > 4$, $s = 0$ is an inflection point while the other two roots are global minimum. The variance in case 1 equals $\sigma^2(\beta) = \frac{1}{4-\beta}$ with $\frac{1}{4-\beta} = \frac{1}{G''_{\beta}(0)} - \frac{1}{\beta}$.

$$G''_{\beta}(s) = \beta - \frac{\beta^2}{4}\mathrm{sech}^2\Big(\frac{\beta s}{2}\Big). \tag{98}$$

■

Sufficient conditions for consistent estimators was studied for the class SERGM Chandrasekhar and Jackson (2014), and by consequence for ERGM. For ERGM, these conditions assumed that the observed network statistics (properly normalised) concentrate around its expected value. First, we show that the expectations-identified and concentrated imply consistency for ERGM, this result was first presented by Chandrasekhar and Jackson (2014) in the more general framework of SERGM. However, our aim is to highlight that the assumptions for having consistency are at the macroscopic level and they are not on the data generating process or behavioral constraints.

Let $\{\mathrm{ERGM}_n = (\Omega_n, 2^{\Omega_n}, P_{\boldsymbol{\theta^n}})\}_{n=2}^{\infty}$ be a sequence of ERGM with probability distributions defined by the same network statistics $\boldsymbol{\Gamma}$ and parameters $\boldsymbol{\theta^n}$.

We say that the sequence ERGM_n is *expectations-identified* with respect to the positive semidefinite matrices $C_n(\tilde{\boldsymbol{\theta}})$ if the following holds. There exists $\gamma > 0$ such that for all $n > 0$ and for all feasible $\tilde{\boldsymbol{\theta}}$, we have that

$$\|C_n(\tilde{\boldsymbol{\theta}})\big(\mathrm{E}_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{\Gamma}) - \mathrm{E}_{\boldsymbol{\theta^n}}(\boldsymbol{\Gamma})\big)\|_2 > \gamma\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta^n}\|_2.$$

We say that the sequence ERGM_n is *concentrated* with respect to the positive semidefinite matrices $C_n(\tilde{\boldsymbol{\theta}})$ if

$$C_n(\hat{\boldsymbol{\theta}}^{\boldsymbol{n}})(\boldsymbol{\Gamma}(\boldsymbol{e_n}) - \mathrm{E}_{\boldsymbol{\theta^n}}(\boldsymbol{\Gamma})) \xrightarrow{P} 0 \quad \text{for } \boldsymbol{\theta^n} \in \Theta.$$

where $\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}$ is such that $\mathrm{E}_{\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}} = \boldsymbol{\Gamma}(\boldsymbol{e_n})$

Result 22. *If a sequence of* ERGM_n *is expectations-identified and concentrated with respect to* $C_n(\tilde{\boldsymbol{\theta}})$*, then the MLE* ($\{\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}\}$) *satisfies that*

$$\hat{\boldsymbol{\theta}}^{\boldsymbol{n}} - \boldsymbol{\theta}^{\boldsymbol{n}} \xrightarrow{P} 0.$$

The reason to our more general definition is evident by noticing that for any pair of parameters $\tilde{\boldsymbol{\theta}}$, $\boldsymbol{\theta}$, we have

$$\left(\mathrm{E}_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{\Gamma}) - \mathrm{E}_{\boldsymbol{\theta}^{\boldsymbol{n}}}(\boldsymbol{\Gamma})\right) = \mathrm{Cov}_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}(\boldsymbol{\Gamma})\left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\boldsymbol{n}}\right). \tag{99}$$

for some $\bar{\boldsymbol{\theta}}^{\boldsymbol{n}} = \lambda\tilde{\boldsymbol{\theta}} + (1-\lambda)\boldsymbol{\theta}^{\boldsymbol{n}}$ and $\lambda \in [0,1]$. Equality (99) follows from the Mean value theorem applied to $\nabla l_{\boldsymbol{e_n}}(\boldsymbol{\theta})$.

As the covariance matrix is positive semidefinite $\mathrm{Cov}_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}(\boldsymbol{\Gamma})$, we can rewrite the covariance matrix as $D_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}} V_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}} D^t_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}$ where $D_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}$ is an orthogonal matrix and $V_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}$ is a diagonal matrix with all entries positive. In Chandrasekhar and Jackson (2014), the definition of expectations-identified and concentrated are given in terms of a sequence of diagonal matrices C_n with entries greater than zero. In this particular case, C_n resembles $V^{-1}_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}$ and $D_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}$ are the identity matrix.

$$D_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}} V^{-1}_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}} D^t_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}\left(\mathrm{E}_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{\Gamma}) - \mathrm{E}_{\boldsymbol{\theta}^{\boldsymbol{n}}}(\boldsymbol{\Gamma})\right) = \left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\boldsymbol{n}}\right). \tag{100}$$

If we let $C_n(\bar{\boldsymbol{\theta}}) = D_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}} V^{-1}_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}} D^t_{\bar{\boldsymbol{\theta}}^{\boldsymbol{n}}}$, then by Formula (99) we have

$$\|C_n(\tilde{\boldsymbol{\theta}})\left(\mathrm{E}_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{\Gamma}) - \mathrm{E}_{\boldsymbol{\theta}^{\boldsymbol{n}}}(\boldsymbol{\Gamma})\right)\|_2 > \gamma\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\boldsymbol{n}}\|_2.$$

for any $\gamma \in (0,1)$. Therefore, all sequence of ERGM are expectations-identified with respect to $C_n(\bar{\boldsymbol{\theta}})$ and the only crucial point in Result 22 is the assumption of concentrated.

Proof of Result 22:

We have for distributions in the exponential family that the MLE $\hat{\boldsymbol{\theta}}^{\boldsymbol{n}} = \boldsymbol{\theta}(\boldsymbol{e_n})$ satisfies $\mathrm{E}_{\hat{\boldsymbol{\theta}}^n}(\boldsymbol{\Gamma}) = \boldsymbol{\Gamma}(\boldsymbol{e_n})$.

Next by the assumption of expectations-identified, we have that

$$\|C_n(\hat{\boldsymbol{\theta}}^{\boldsymbol{n}})\big(\mathrm{E}_{\hat{\boldsymbol{\theta}}^n}(\boldsymbol{\Gamma}) - \mathrm{E}_{\boldsymbol{\theta}}(\boldsymbol{\Gamma})\big)\|_2 > \gamma\|\hat{\boldsymbol{\theta}}^{\boldsymbol{n}} - \boldsymbol{\theta}\|_2$$

Finally, we know that the left side converges to zero and the right side is greater or equal than zero.

Thus, $\|\hat{\boldsymbol{\theta}}^{\boldsymbol{n}} - \boldsymbol{\theta}\|_2$ converges to zero in probability.

■

Result 23. *If the assumptions of Result 20 are satisfied, then the sequence of C-ERGM are expectations-identified and concentrated.*

The proof is an inmediate consequence of Result 25.

7.4 Similarities between finite exponential random graph and exponential random graph with community structure

Proof of Result 19:

By Theorem 8, we have that the joint distribution of the variables $\boldsymbol{e_n}$ is

$$P_{\boldsymbol{\theta}^n}(\boldsymbol{e_n}) = \frac{1}{c_n(\boldsymbol{\theta^n})} \exp\left(\sum_{k=1}^{p+q} \theta_k^n \Gamma_k(\boldsymbol{e_n})\right). \tag{101}$$

Notice that for $k \leq p$, we have that $\Gamma_k(\boldsymbol{e_n}) = \sum_{i=1}^{u_n} \Gamma_k(\boldsymbol{e}_{\mathcal{U}_i^n})$. While for $p < k \leq q$, we have that $\Gamma_k(\boldsymbol{e_n}) = \sum_{i=1}^{w_n} \Gamma_k(\boldsymbol{e}_{W_i})$.

Thus, we can decompose the joint distribution as follows

$$P(\boldsymbol{e_n}) = \frac{1}{c_n(\boldsymbol{\theta^n})} \prod_{i=1}^{n_u} Q^{u_i}(\boldsymbol{e}_{\mathcal{U}_i^n}) \prod_{i=1}^{n_w} Q^{w_i}(\boldsymbol{e}_{W_i})$$

where $Q^{u_i}(\boldsymbol{e}_{\mathcal{U}_i^n}) = \exp\left(\sum_{k=1}^{p} \theta_k^n \Gamma_k(\boldsymbol{e}_{\mathcal{U}_i^n})\right)$ and $Q^{w_i}(\boldsymbol{e}_{W_i}) = \exp\left(\sum_{k=p+1}^{p+q} \theta_k^n \Gamma_k(\boldsymbol{e}_{W_i})\right)$.

Finally, by observing that

$$1 = \frac{1}{c_n(\boldsymbol{\theta^n})} \prod_{i=1}^{n_u} \overbrace{\int Q^{u_i}(\boldsymbol{e}_{\mathcal{U}_i^n}) dP(\boldsymbol{e}_{\mathcal{U}_i^n})}^{A_i} \prod_{i=1}^{n_w} \overbrace{\int Q^{w_i}(\boldsymbol{e}_{W_i}) dP(\boldsymbol{e}_{W_i})}^{B_i}$$

and defining $P_{\boldsymbol{\theta_u^n}}(\boldsymbol{e}_{\mathcal{U}_i^n}) = \frac{Q^{u_i}(\boldsymbol{e}_{\mathcal{U}_i^n})}{A_i}$ and $P_{\boldsymbol{\theta_w^n}}(\boldsymbol{e}_{W_i}) = \frac{Q^{w_i}(\boldsymbol{e}_{W_i})}{B_i}$, we have that

$$P(\boldsymbol{e_n}) = \prod_{i=1}^{n_u} P_{\boldsymbol{\theta_u^n}}(\boldsymbol{e}_{\mathcal{U}_i^n}) \prod_{i=1}^{n_w} P_{\boldsymbol{\theta_w^n}}(\boldsymbol{e}_{W_i})$$

■

Proof of Result 20: The proof for $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{\boldsymbol{n}} \to \boldsymbol{\theta_u}$ leans on showing that the assumptions of Theorem 6.3.1 in the script "Mathematical Statistics" by Sara van Geer holds. This follows from three points: i) the independence of the variables $\{\boldsymbol{e}_{\mathcal{U}_i^n}\}$, ii) $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{\boldsymbol{n}}$ is an M-estimator, iii) $\mathrm{E}_{\boldsymbol{\theta}}(\sup_{\tilde{\boldsymbol{\theta}}} |\log(p_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{e}_{\mathcal{U}})|)) < \infty$ and iv) the true parameter is a well-separated point, i.e. for all $\delta > 0$ $\sup\{\mathrm{E}_{\boldsymbol{\theta}}(\log(p_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{e}_{\mathcal{U}}))) : \tilde{\boldsymbol{\theta}} \in \Theta \ \ ||\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}||_2 > \delta\} < \mathrm{E}_{\boldsymbol{\theta}}(\log(p_{\boldsymbol{\theta}}(\boldsymbol{e}_{\mathcal{U}})))$.

Point i) follows from Result 24 and ii) is true by construction. For iii), it is sufficient to notice that the supremum is taken in a compact set, hence it finite and the integral is the sum of finite numbers. For iv), we have that

$$\sup\{\mathrm{E}_{\boldsymbol{\theta}}(\log(p_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{e}_{\mathcal{U}}))) : \tilde{\boldsymbol{\theta}} \in \Theta \ \ ||\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}||_2 > \delta\} \le \sup\{\mathrm{E}_{\boldsymbol{\theta}}(\log(p_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{e}_{\mathcal{U}}))) : \tilde{\boldsymbol{\theta}} \in \Theta \ \ ||\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}||_2 \ge \delta\}.$$

The set $\{\tilde{\boldsymbol{\theta}} \in \Theta \ \ ||\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}||_2 \ge \delta\}$ is the intersection of a compact and a close set, hence it is compact and its supremum equals $\mathrm{E}_{\boldsymbol{\theta}}(\log(p_{\bar{\boldsymbol{\theta}}}(\boldsymbol{e}_{\mathcal{U}})))$ for some $\bar{\boldsymbol{\theta}}$ in the set. By assumption, $\boldsymbol{\theta}$ is the unique maximiser and thus $\mathrm{E}_{\boldsymbol{\theta}}(\log(p_{\bar{\boldsymbol{\theta}}}(\boldsymbol{e}_{\mathcal{U}}))) < \mathrm{E}_{\boldsymbol{\theta}}(\log(p_{\boldsymbol{\theta}}(\boldsymbol{e}_{\mathcal{U}})))$.

The proof for $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}^{\boldsymbol{n}} \to \boldsymbol{\theta_w}$ follows a similar argument.

Next, we show the asymptotic normality in the Limit (83):

$$\frac{\Gamma_k(\boldsymbol{e_n}) - \mathrm{E}_{\boldsymbol{\theta}}(\Gamma_k)}{\sqrt{\mathrm{Var}_{\hat{\boldsymbol{\theta}}^n}(\Gamma_k)}} = \overbrace{\frac{\sqrt{\mathrm{Var}_{\boldsymbol{\theta}}(\Gamma_k)}}{\sqrt{\mathrm{Var}_{\hat{\boldsymbol{\theta}}^n}(\Gamma_k)}}}^{(A)} \overbrace{\frac{\Gamma_k(\boldsymbol{e_n}) - \mathrm{E}_{\boldsymbol{\theta}}(\Gamma_k)}{\sqrt{\mathrm{Var}_{\boldsymbol{\theta}}(\Gamma_k)}}}^{(B)} \xrightarrow{D} \mathcal{N}(0,1) \tag{102}$$

The convergence follows from the Slutsky's theorem: (A) converges to 1 is shown in Result 25 and (B) converges to the normal distribution is shown in Theorem 9.

■

Result 24. *1. If U, U' are two communities and the cardinality of $U \cap V(\boldsymbol{e_n})$ is equal to the cardinality of $U' \cap V(\boldsymbol{e_n})$ and it is greater than zero, then $\boldsymbol{e}_{\mathcal{U}_n}$ and $\boldsymbol{e}_{\mathcal{U}'_n}$ are independent and identically distributed. $\mathcal{U}_n$ is the set of pair of agents in $V(\boldsymbol{e_n})) \cap U$, i.e. $\big((U \cap V(\boldsymbol{e_n})) \times (U \cap V(\boldsymbol{e_n}))\big) \setminus \{ii\}_{i \in U}$.*

Let $W = W(U, U') = U \times U' \bigcup U' \times U$ denotes the set of pairs of agents across two communities U, U'; and W' a different set of pairs of agents across two communities U_1 and U'_1.

If the cardinality of U (U') is equal to the cardinality of U_1 (U'_1) then e_W and $e_{W'}$ are independent and identically distributed.

On the other hand, if the cardinality of $\boldsymbol{e}_{\mathcal{U}'_n}$ is greater than the cardinality of $\boldsymbol{e}_{\mathcal{U}_n}$ and $\boldsymbol{e}_{\mathcal{U}''_n} \subseteq \boldsymbol{e}_{\mathcal{U}'_n}$ with $|\boldsymbol{e}_{\mathcal{U}''_n}| = |\boldsymbol{e}_{\mathcal{U}_n}|$ it does not follow that marginal distribution of $\boldsymbol{e}_{\mathcal{U}''_n}$ equals to the marginal distribution of $\boldsymbol{e}_{\mathcal{U}_n}$.

Although, the conditional $\boldsymbol{e}_{\mathcal{U}''_n} | \boldsymbol{e}_{\mathcal{U}'_n \setminus \mathcal{U}''_n} = \mathbf{0}$ has the same distribution as $\boldsymbol{e}_{\mathcal{U}_n}$; the interpretation is quite different. While the conditional represents the updating of relations between the agents $V(\boldsymbol{e}_{\mathcal{U}''_n})$ when all the other possible relations are set to non-existence, the unconditional distribution does not restrict the existence or non-existence of relations.

Result 25. *If the assumptions of Result 20 are satisfied, then*

$$\frac{\sqrt{\mathrm{Var}_{\boldsymbol{\theta}}(\Gamma_k(\boldsymbol{e_n}))}}{\sqrt{\mathrm{Var}_{\hat{\boldsymbol{\theta}}^n}(\Gamma_k(\boldsymbol{e_n}))}} \to 1$$

Proof of Result 25:

For $k \leq p$, we have that $\mathrm{Var}_{\tilde{\boldsymbol{\theta}}}(\Gamma_k(\boldsymbol{e_n})) = n_u \mathrm{Var}_{\tilde{\boldsymbol{\theta}}}(\Gamma_k(\boldsymbol{e}_{\mathcal{U}}))$. Therefore, we only have to show that $\mathrm{Var}_{\hat{\boldsymbol{\theta}}^n}(\Gamma_k(\boldsymbol{e}_{\mathcal{U}})) \to \mathrm{Var}_{\boldsymbol{\theta}}(\Gamma_k(\boldsymbol{e}_{\mathcal{U}}))$. This follows from the facts that a) $\hat{\boldsymbol{\theta}}^{\boldsymbol{n}}$ converge to $\boldsymbol{\theta}$, $\mathrm{Var}_{\tilde{\boldsymbol{\theta}}}(\Gamma_k(\boldsymbol{e}_{\mathcal{U}}))$ is a continuous function and that convergence in probability is preserved under continuous transformations. For $k \in \{p+1, .., p+q\}$ a similar argument follows.

■

Result 26. *Let $P_{\boldsymbol{\theta}^n}$ be a sequence of ERGM satisfying Assumptions 5-7 and let us assume that the sequence is δ-sparse with $\delta > 2$, then*

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}^n}(\Gamma_{1,W}(\boldsymbol{e_n}) > \epsilon) = 0 \ \ \forall \epsilon > 0$$

The assumption of δ-sparse with $\delta > 2$ is necessary if the model includes triangles or any other structure H_k with $|V(H_k)| \geq 2$ (see Theorem 2 in Schweinberger and Handcock (2015)).

Proof of Result 26:

Since, $\Gamma_{1,W}(\boldsymbol{e_n})$ is a nonnegative function by the Markov inequality, we have that for all $\epsilon > 0$

$$P_{\boldsymbol{\theta}^n}(\Gamma_{1,W}(\boldsymbol{e_n}) \geq 0) \leq \frac{\mathrm{E}_{\boldsymbol{\theta}^n}(\Gamma_{1,W}(\boldsymbol{e_n}))}{\epsilon}$$

Now, a crude upper bound for $\mathrm{E}_{\boldsymbol{\theta}^n}(\Gamma_{1,W}(\boldsymbol{e_n}))$ is

$$\mathrm{E}_{\boldsymbol{\theta}^n}(\Gamma_{1,W}(\boldsymbol{e_n})) = \sum_{l=1}^{n_w} \sum_{e_{ij} \in \boldsymbol{e}_{W_l}} \mathrm{E}_{\boldsymbol{\theta}^n}(e_{ij}) \leq cn^{2-\delta}$$

Thereby, $\mathrm{E}_{\boldsymbol{\theta}^n}(\Gamma_{1,W}(\boldsymbol{e_n})) \to 0$

■

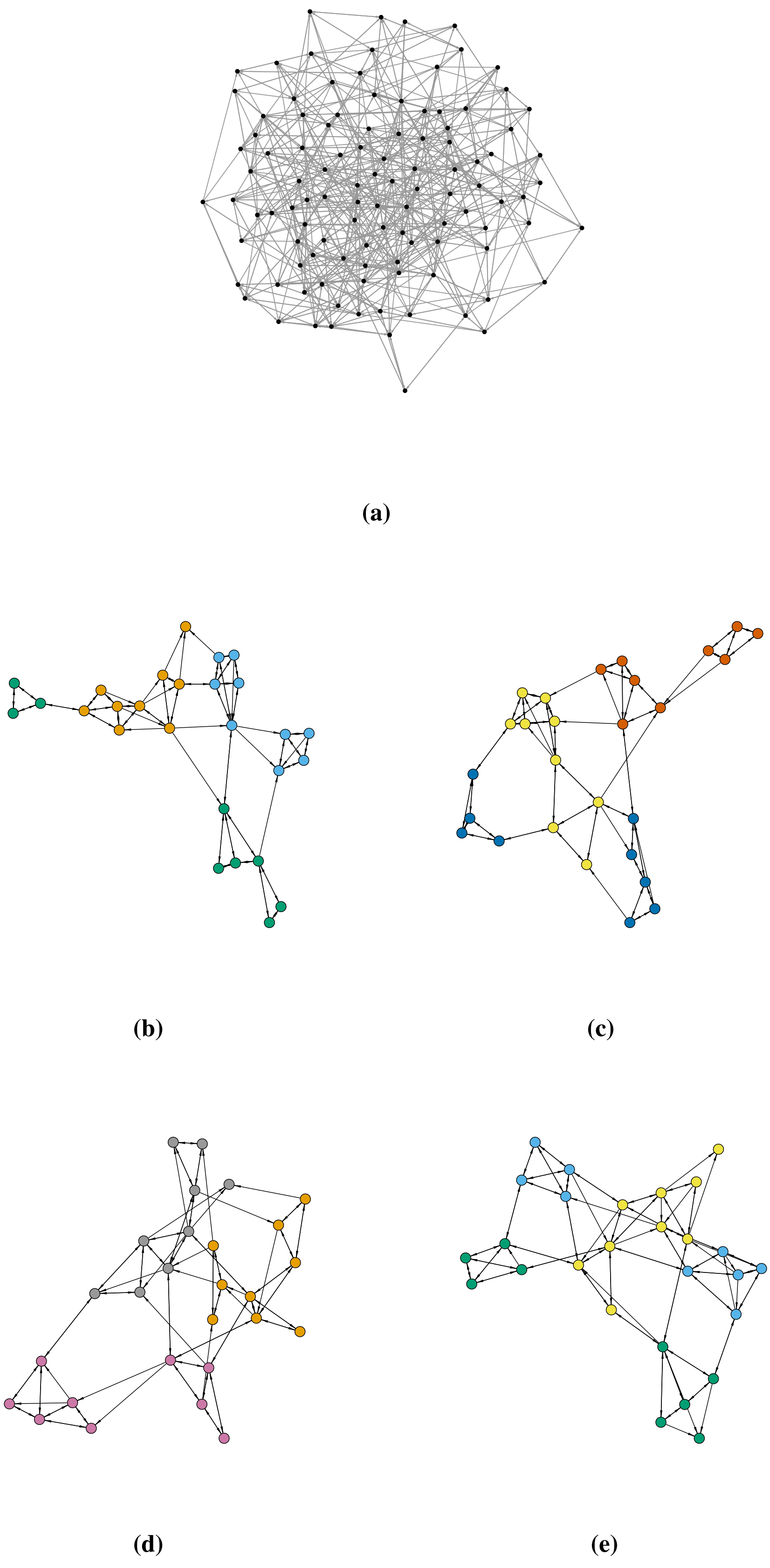


Figure 21: **(a)** illustrates a network with 12 communities and each community has 9 agents, relational variables within a community are dependent but independent otherwise. **(b)-(e)** illustrate four different subnetworks, each subnetwork is defined by the relations between agents in three communities. The colour of a node represents the community of the agent.

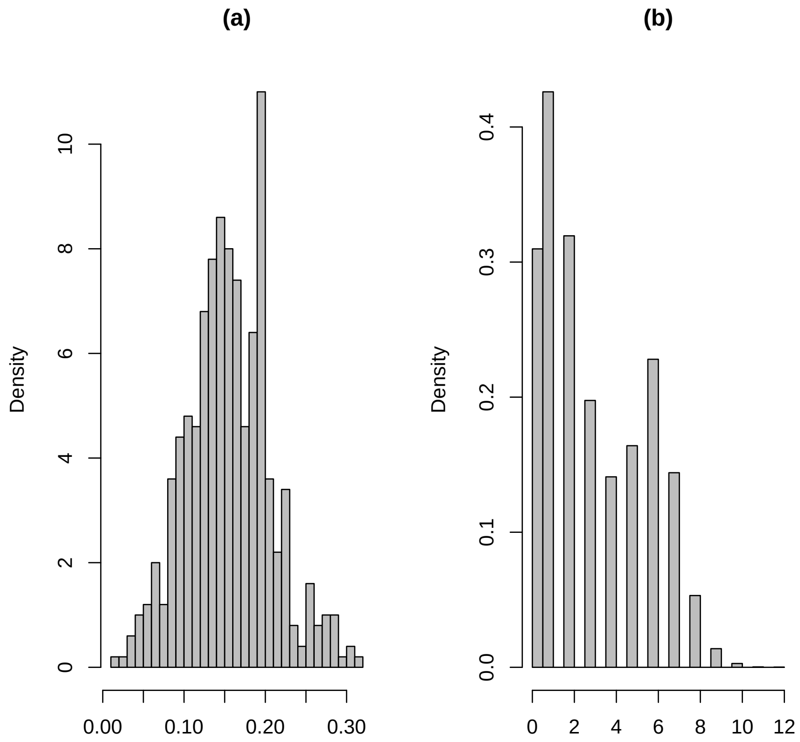
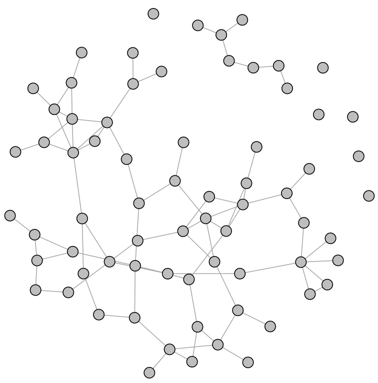
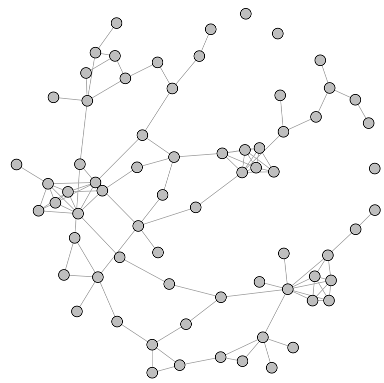


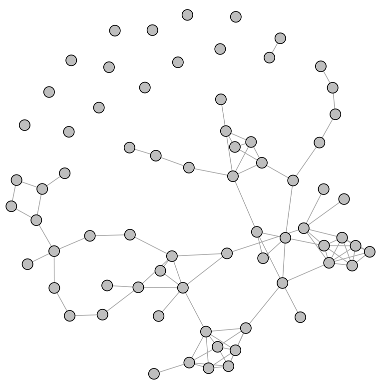
Figure 22: Histogram **(a)** shows the distribution of the number of isolated agents in the network and histogram **(b)** shows the degree distribution.



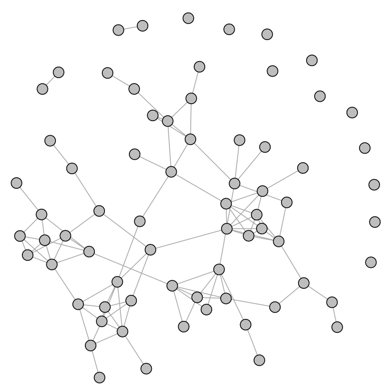
(a)



(b)



(c)



(d)

Figure 23: (a)-(d) illustrate four different networks sampled from the same C-ERGM, in all cases, we observe a large connected component, several isolated agents and cluster of agents.

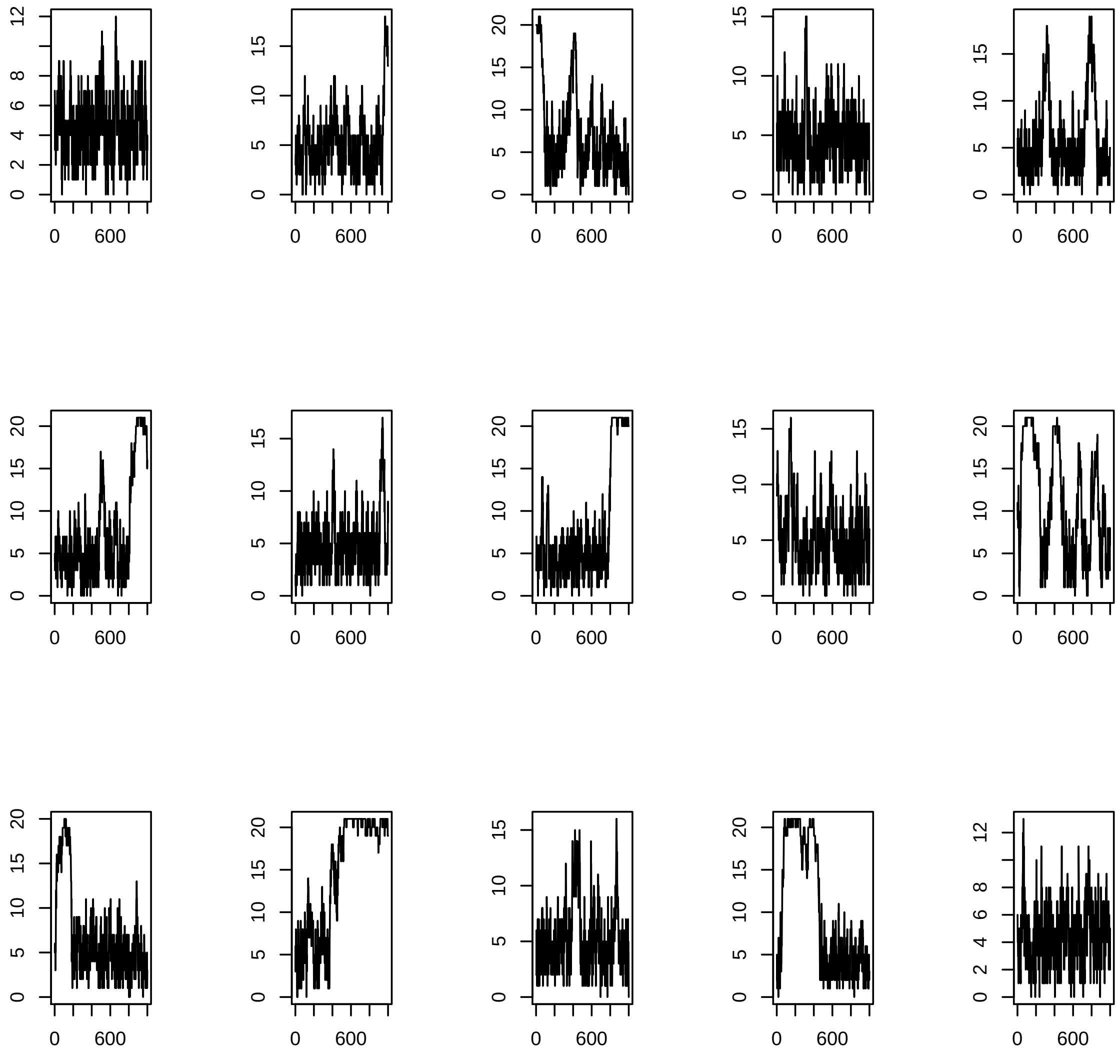


Figure 24: Each plot shows the evolution of the number of relations within a particular community. Although it is possible to observe communities where almost all agents are connected, at some point the number of relations will decrease dramatically.

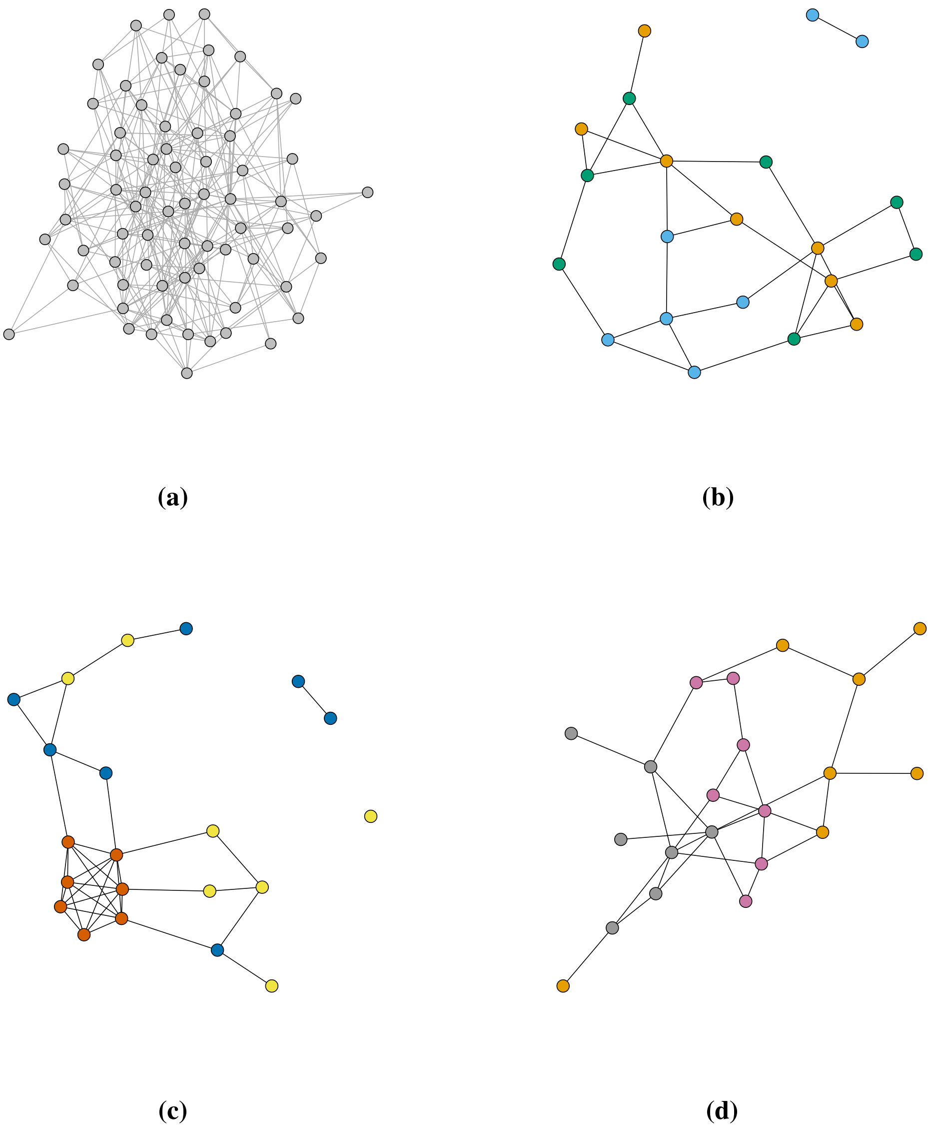


Figure 25: **(a)** illustrates a network sampled from a C-ERGM with 10 communities and each community consists of 7 agents. **(c)**-**(d)** illustrate three different subnetworks, each subnetwork is defined by the relations between agents in three communities. The colour of a node represents the community of the agent.

8 Summary and Outlook

In this dissertation, we address different issues researcher faced when performing statistical analysis on network data. In Chapter 2, we construct a class of network models, power random graph models (PRGM), by introducing assumptions at the agent-agent interaction level. The assumptions at the microscopic level are defined using the concept of q-product and q-conditional independence. We show that the q-product can be interpreted as an idiosyncratic correlation or as introducing interaction terms between the social mechanisms defining the model. When $q = 1$, the q-product and q-conditional independence reduce to the classical product and conditional independence. Further, we show that PRGM has an entropy foundation using Tsallis entropy. PRGM enriches network models by moving beyond network models based on Boltzmann-Shannon entropy. In Section 2.5, we define a subclass of PRGM, termed q-Markov random graph models, which have as a particular case the well-known Markov random graph models. Using simulations, we show that the bimodal distributions documented in some Markov random graph models are the result of two opposite social mechanisms in the formation of relations between agents, and the absence of an idiosyncratic correlation. By increasing the idiosyncratic correlation, we show that the probability mass on the complete and/or null graph decreases. Nonetheless, the probability distribution over the number of relations is not well approximated by a Gaussian distribution, and thus p-values cannot be justified on the ground of Gaussian approximations.

In Chapter 4, we construct a network model, finite exponential random graph model (fERGM), which allows constructing consistent estimators as the number of networks tends to infinity and without fixing the size of the networks. We show how the fERGM can be applied for testing if the number of agents in a network influences the tendency of agents to reciprocate relations or the tendency to become a friend of a friend. Although analysing multiple networks is not a new topic, we show that one of the most used approaches for analysing multiple networks lacks on fundamental statistical properties, e.g. estimators are not consistent.

In Chapter 6, we review the exponential random graph model. We introduce the ERGM with a community structure, and we present conditions under which the central limit theorem holds as the number of communities in the network tends to infinity. In Section 6.4, we review the underlying assumption of subclasses of ERGM. Next, we show that bimodal distribution frequently observed in ERGM are the result of its underlying mathematical framework, and not necessarily a result of a misspecified model. Contrary to previous approaches dismissing model with a bimodal distribution, our results suggest that bimodality may occur at the community level while it may be absent at the macroscopic level, and thus it may suggest a small sample size problem.

Despite the advantage shown in the PRGM and fERGM over ERGM, some problems lie ahead. First, the computational complexity to estimate power law random graph models makes it almost impossible for researchers to apply PRGM to large networks, and thus further research on developing estimation methods is required. Similarly, fERGM suffers from computational problems. Second, since it is unknown if CLT holds for network data, it is necessary to develop new statistical frameworks for data analysis. Additionally, practitioners need to be critical about existing methodologies for network data interpretation which make use of (weak-)independence assumptions.

Curriculum Vitae

Personal details

Abel Camacho Guardian

Date of birth: 25 October 1986

Education

08/14 – 02/18	Doctoral Student at the University of Zurich, Department of Business Administration, Chair of Marketing and Market Research
Since 03/13	Member of the University Research Priority Program on “Social Networks", University of Zurich
09/10 – 05/12	Master of Science in Mathematics, Eidgenossische Technische Hochschule Zurich, ETH Zurich.
09/05 – 08/09	Bachelor of Science in Mathematics, National Autonomous University of Mexico, UNAM.

Professional experience

Since 03/13	Teaching assistant at the Chair of Marketing and Market Research, University of Zurich.
11/12 – 02/13	Research Assistant, University of St Gallen, St Gallen.
06/12 – 09/12	Internship as Business Analyst, Partners Group, Zug.
08/08 – 12/08	Teaching assistant, UNAM, Mexico City.